

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:31:35 ; Search time 61.65 Seconds

(without alignments)
518,460 Million cell updates/sec

Title: US-09-147-955-6

Perfect score: 2407

Sequence: 1 MSRAHVLLATFPAGGHINPA.....EEDGSSVNNKATVLEDEVG 461

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP. archaea:*
2: SP. bacteria:*
3: SP. fungi:*
4: SP. human:*
5: SP. invertebrate:*
6: SP. mammal:*
7: SP. mhc:*
8: SP. organelle:*
9: SP. phage:*
10: SP. plant:*
11: SP. rodent:*
12: SP. virus:*
13: SP. vertebrate:*
14: SP. unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2407	100.0	461	10	092R25	092R25 verbena x h
2	1646.5	68.4	460	10	092R27	092R27 perilla fru
3	1583.5	65.8	443	10	092R26	092R26 perilla fru
4	1054.5	43.8	456	10	032R20	032R20 arabidopsis
5	1052	43.7	467	10	P93709	P93709 nicotiana t
6	1022.5	42.5	474	10	004930	004930 arabidopsis
7	1003.5	41.7	519	10	092VY2	092VY2 arabidopsis
8	988	41.0	455	10	092VY5	092VY5 arabidopsis
9	935.5	38.9	458	10	023406	023406 arabidopsis
10	779	32.4	449	10	022822	022822 arabidopsis
11	740	30.7	460	10	048676	048676 arabidopsis
12	730	30.3	449	10	022820	022820 arabidopsis
13	711.5	29.6	479	10	023401	023401 arabidopsis
14	697	29.0	456	10	022182	022182 arabidopsis
15	682	28.3	438	10	022183	022183 arabidopsis
16	682	28.3	484	10	023400	023400 arabidopsis
17	656	27.3	453	10	022186	022186 arabidopsis
18	631	26.2	475	10	023402	023402 arabidopsis
19	554	23.0	481	10	092WJ3	092WJ3 arabidopsis

20	535	22.2	482	10	092UV0	092UV0 arabidopsis
21	504.5	21.0	455	10	064733	064733 arabidopsis
22	485.5	20.2	495	10	092Q95	092Q95 arabidopsis
23	483	20.1	440	10	064732	064732 arabidopsis
24	479	19.9	462	10	004622	004622 arabidopsis
25	479	19.9	495	10	092Q94	092Q94 arabidopsis
26	478	19.9	491	10	092Q99	092Q99 arabidopsis
27	475.5	19.8	496	10	092Q97	092Q97 arabidopsis
28	471.5	19.6	496	10	092Q96	092Q96 arabidopsis
29	466.5	19.4	484	10	092Q64	092Q64 arabidopsis
30	461.5	19.2	496	10	092Q98	092Q98 arabidopsis
31	457	19.0	476	10	P93364	P93364 nicotiana t
32	451.5	18.8	454	10	092X16	092X16 forsythia x
33	447	18.6	477	10	092W05	092W05 vigna mungo
34	445	18.5	476	10	P93365	P93365 nicotiana t
35	431	17.9	347	10	049042	049042 pisum sativ
36	429	17.8	466	10	043526	043526 lycopersico
37	428	17.8	478	10	049492	049492 arabidopsis
38	425	17.7	481	10	082381	082381 arabidopsis
39	424.5	17.6	474	10	082382	082382 arabidopsis
40	418	17.4	455	10	092W52	092W52 vigna mungo
41	410	17.0	460	10	092Q63	092Q63 arabidopsis
42	405.5	16.8	488	10	P93789	P93789 solanum tub
43	403.5	16.8	447	10	004114	004114 perilla fru
44	397	16.5	456	10	022304	022304 vitis vinif
45	394	16.4	452	10	022303	022303 vitis vinif

ALIGNMENTS

RESULT	ID	Q92R25	PRELIMINARY:	PRT:	461 AA.
Q92R25	AC	Q92R25			
Q92R25	DT	01-MAY-1999 (TREMBLREL. 10, Created)			
Q92R25	DT	01-MAY-1999 (TREMBLREL. 10, Last sequence update)			
Q92R25	DT	01-NOV-1999 (TREMBLREL. 12, Last annotation update)			
Q92R25	DE	UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.			
Q92R25	GN	HGT9.			
Q92R25	OS	Verbena x hybrida.			
Q92R25	OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Q92R25	OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
Q92R25	OC	core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;			
Q92R25	OC	Verbena.			
Q92R25	RN	[1]			
Q92R25	RP	SEQUENCE FROM N.A.			
Q92R25	RC	TISSUE-PETAL.			
Q92R25	RX	MEDLINE: 99167509.			
Q92R25	RA	YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,			
Q92R25	RA	KUSUMI T., SAITO K.;			
Q92R25	RT	"Molecular cloning and biochemical characterization of a novel			
Q92R25	RT	anthocyanin 5-O-glucosyltransferase by mRNA differential display for			
Q92R25	RT	plant forms regarding anthocyanin."			
Q92R25	RT	J Biol Chem 274:7405-7411(1999).			
Q92R25	DR	EMBL: AB013588; BAA36423.1; -			
Q92R25	DR	PROSITE: PS00375; UDPGT: 1.			
Q92R25	KW	Transferase.			
Q92R25	SQ	SEQUENCE 461 AA; 51347 MW; 6DCBC3E6 CRC32;			
Q92R25	Query Match	100.0%; Score 2407; DB 10; Length 461;			
Q92R25	Best Local Similarity	100.0%; Pred. No. 3e-178;			
Q92R25	Matches	461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Q92R25	QY	1 MSRAHVLLATFPAGGHINPA...QFAKRLANDIVFTTSYAWRRSRTPAGSNGILNFV 60			
Q92R25	DB	1 MSRAHVLLATFPAGGHINPA...QFAKRLANDIVFTTSYAWRRSRTPAGSNGILNFV 60			
Q92R25	QY	61 SPSDGTGDLGPDGDKGNVSEKSRGIRKLSOTLAANNVDKSSRTFVVYVSHLFAANA 120			
Q92R25	DB	61 SPSDGTGDLGPDGDKGNVSEKSRGIRKLSOTLAANNVDKSSRTFVVYVSHLFAANA 120			

[illegible]

OY 235 IPSAFGLDCKPDSRSCGCDLFEKGSNDDDCLSTINRSGSVVSVSGSPV - WTTXSKQK 293
 241 VSSS - BGKT-----DLFK--SSDEDTYKWKDCKLEREVITISLGTADADLPKKNH 288
 OY 294 EIANGLLDGCRPLMYVVRNGBEVLISCMELKTR--VGKIVSGSOLEVLTPLSGCF 350
 DB 289 ALTGATVATNRPLMLVTRKBNPEKKNNFETLRGSDRGVLWGCQOTVLAHCAVWGCF 348
 OY 351 VTRCGMNSTLESTISFCVPMVAEPFQDGTNAKLMEDEVTRFVGRVARNEGSGVADDEIR 410
 DB 349 VTRCGMNSTLESTISGCVPMVAEPFQDGTNAKLMEDEVTRFVGRVARNEGSGVADDEIR 407
 OY 411 RCIEEVDGGEKSRKTRRESNGKWKDLARAMEGDSVNNKAVLDE 457
 DB 408 RCLEKVSNGSGEAEEMKENNERKMAVADAAEGGPSDLNLKGFVDE 454

RESULT 5
 P93709 PRELIMINARY: PRG: 467 AA.
 AC P93709:
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
 DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
 DE GLUCOSYL TRANSFERASE.
 GN JET.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 RN Nicotiana.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BF-2; TISSUE-SUSPENSION CULTURE;
 RC KOUTIMA R., HASHIZUME K., IMANISHI S., NAKAMURA K.;
 RA Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AB006623; BAA19155.1; -
 DR MENDEL: 9421; Nicta:1525;9421.
 DR PROSITE: PS00375; UDPGT: 1.
 DR PFM: PFM0201; UDPGT: 1.
 KM Transferrase.
 SO SEQUENCE 467 AA; 52989 MW; 3726EB31 CRC32;

Query Match 43.7%; Score 1052; DB 10; Length 467;
 Best Local Similarity 45.1%; Pred. No. 1,9e-73;
 Matches 211; Conservative 93; Mismatches 140; Indels 24; Gaps 11.

OY 2 SRNAVLTATPAGCHINPALQPAKRLANADIOYTFYISVVARMSSTAAGSGLNIPVS 61
 DB 6 NECHVLLALTEPGGCHINPISLOFSKRLINGVKYLSSLSLSAFNRI-KNLPRIEGL-TFAP 63
 OY 62 FSDSDVDLGLPG--DDCKNMYSEMKSRCJIKALSDTLAANNVDOKSK--TFVYVSHLPA 117
 DB 64 FSDSDYDGNFGSDYDHLFNSAIKSHGSEFI----ANLISKAKNGYPTPTVITYITLMD 118
 OY 118 MAATVAEPEFLREKALIMTEPATVIOFEYFNGYSD---ETDAGSDAHLPGGLPIALOR 174
 DB 119 WAGSVAKRLHISTLMLEIOPATVFDYIYKTFNANFKNTYSODQITELP-GLPSLSS 177
 OY 175 DLPSFLIPSTHERFSL--MKKELETLEGEKPKVLVNSFDALPDALKADRYEMIAIG 232
 DB 178 DLPSPFVDDVKSNDMAVSEIKROIETLNSSEKPRILVTFPALELALNLRPLKANVTYWGJ 237
 OY 233 PLIPSAFLDCKPDSRSCGCDLFEKGSNDDDCLSTINRSGSVVSVSGSPVTFKTSQK 292
 DB 238 PLIPSPSLDCKPDSRSCGCDLFEKGSNDDDCLSTINRSGSVVSVSGSPVTFKTSQK 294
 OY 293 EIAFARGLLDGGREFLVVVRV---NEGEVILISCMELKRVGKIVSGSOLEVLTPLSLGC 349
 DB 295 EETISGGLCKGRPLMVVIRETLNGEKPREKLTCKDDELEKIGIRIVMCSQMEVLRKSHSVGC 354

Db 503 6 503

RESULT 8

092YV5 PRELIMINARY: PRT: 455 AA.

AC 092YV5: 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE T25N20.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta:
 euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: Rosidae: eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SHINN P., BUEHLER E., DEMAR K., PENG J., KIM C., LI Y., SUN H.,
 RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,
 RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
 RA ECKER J.R.,
 RT "Genomic sequence for Arabidopsis thaliana BAC T25N20."
 RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC005106; AAC80596.1; -
 DR PROSITE; PS00375; UDPGT; 1.
 SQ SEQUENCE 455 AA; 51190 MW; F9BAC33E CRC32;

Query Match 41.0%; Score 988; DB 10; Length 455;

Best Local Similarity 45.3%; Pred. No. 1.6e-68;
 Matches 213; Conservative 84; Mismatches 139; Indels 34; Gaps 12;

QY 1 MSRAHVLATFPAGCHINPALQFAKRL-ANADIQVTFYTSVYAMRMSRTAAGSGLINF 59
 DB 1 MAQHFLLVTFPAGCHINPALQFAKRL-ANADIQVTFYTSVYAMRMSRTAAGSGLINF 59
 QY 60 VSPEDGYDDG-LQPGDGKMYSEMKSRGIKALSDTLAANNVDOKSKITFYVSHLEAF 118
 DB 61 LTFSDGDDGVYSTNDVONRLVHFRNDKALSDITFAN-ONGDSVSCLTITLIPNW 118
 QY 119 AAKVAREFHLSALMIEPATVLDIFYFYNGYSDEIDAGSDAHLPGGLVLAQRDLP 178
 DB 119 VPKVAREFHLPVSHLMQPAFAFDIYNN-SSGNNSVEFEFNLPSLEIRDLP 170
 QY 179 FLFPLS-THREFRSLMKEKLETEGEKPKVLYNSFDALPDAKAIKYEIAIGPLIPS 237
 DB 171 FLSPSNTNKAQAQVAYQELMDPELKEESNPKILVNTFDSLEPEFLTAIPTEKVAAGPLPA 230
 QY 238 AFLDGKPSDRSFGDLFEKGSNDQCLEMLSTNPRSSVYVYSGSEFNTKSMOMEIAR 297
 DB 231 EIFPGSES-----GKDLSDHGSSTYTL-WLDSKTESSVIYVSFGTVELSKQIEELAR 284
 QY 298 GLDQGGPELVAV--RYN-----EGE-----EVLISCMEELKRVKITYSMCSOLEVLT 345
 DB 285 ALLEGGRPELVAVITDKLNRPAKIEGSEETEIEKIAGFHRELEEGMTVSMCSOLEVLT 344
 QY 346 SLGCFVTHCGMNSTLESISFGVPYVAFPOWFDQGTNAKLMEVDVMTGVRVANEESVVD 405
 DB 345 AIGCFELTHCGMNSTLESISLVGYVYVAFPMWSDQPANAKULEIWMGTGVRVANESEGLVE 403
 QY 406 GDEIRRCIEEVMGGEKSRKRLRESAGKWKDLARKAMEDGSSVNNLAKYFL 455
 DB 404 RGEIMRCLAEVMEA--KSYELRENNAEKMRKLATEAGRGSGSDKWEAFV 451

RESULT 9

023406 PRELIMINARY: PRT: 458 AA.

AC 023406: 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE SIMILAR TO INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Eubryophyta: Tracheophyta:
 euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: Rosidae: eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., STEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRY N.,
 RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
 RA PUIGDOMENICH P., HATZIOPOULOS P., OBERMAIER B., DUSTERHOFT A.,
 RA JONES J., PALME K., ANSORGE W., DEISENY M., BANCROFT I., MENES H.W.,
 RA SCHUELLER C., CHALMARTIS N.,
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; Z97339; CAB10333.1; -
 DR MENDEL; 26727; Arabid-2543;26727.
 DR PROSITE; PS00375; UDPGT; 1.
 DR PFAM; PF00201; UDPGT; 1.
 DR Transferase.
 SQ SEQUENCE 458 AA; 52005 MW; 67943A2F CRC32;

Query Match 38.9%; Score 935.5; DB 10; Length 458;

Best Local Similarity 42.7%; Pred. No. 1.8e-64;
 Matches 206; Conservative 82; Mismatches 126; Indels 69; Gaps 16;

QY 5 HVLATFPAGCHINPALQFAKRLAN-ADIQVTFYTSVYAM-RMSRTAAGSGLINF 61
 DB 13 HFLVTFPAGCHINPALSLEAKRLAGTSGARVFAASISAYRRMRFPSTENPETLI-FAT 71
 QY 62 FSDGDDQGLQPG-----DDGKNYSEMKSRGIKALSDTLAANNVDOKSKITFYVYS 113
 DB 72 YSDGDDQFSSKASVSDKSDQDANGNFMSEMRKRGKLELLEIDNR--KQNPPTCVYTT 129
 QY 114 HLFMAKAVAREFHLSALMIEPATVLDIFYFYNGYSDEI-----DAGSDAHLPGGLP 169
 DB 130 ILTLTVAVELA-----LFSIFYFYNGYSDEIASEMANPPSSIKLP-SLP 172
 QY 170 VLAQRDLPFLPSTHERF-RSLMKEKLETEGEKPKVLYNSFDALPDAKAI-DKYE 227
 DB 173 LLYVADIPSPFVSSNVYAFLLPAFREGIDSKELINPKILINTFOLELEPAMSSVPODFK 232
 QY 228 MIAIGPLPSAFLDGKPSDRSFGDLFEKGSNDQCLEMLSTNPRSSVYVYSGSEFVNT 287
 DB 223 IVPVGPLL-----TLRTDFSSRG-----EYIMLDTKADSSVLYVSGETLAVL 275
 QY 288 TRKSMOEFARGILDCGRPLVAV-----RVNAGEVYISCMELKRVKITYSMCS 337
 DB 276 SKKQVLELCKALIOSRRPLVAVITDKSYRKKEDEKEDKEDICSSRRELEIDGMYVSKCD 335
 QY 338 QLEVLTPLSCLCFVTHCGMNSTLESISFGVPYVAFPOWFDQGTNAKLMEVDVMTGVRV 395
 DB 336 QFVLYNHRSLGCFVTHCGMNSTLESISLVGYVYVAFPMWSDQPANAKULEIWMGTGVRV 395
 QY 396 RANEGG-SVYDGDRIIRCIEEVMGGEKSRKRLRESAGKWKDLARKAMEDGSSVNNLAKY 454
 DB 396 KKEBGVYVYVDSERIRRCIEVME--DKAEFFGNATRKQDLAALAVYDEGSSFPNHLKAF 453
 QY 455 LDE 457
 DB 454 VDE 456

RESULT 10

022822 PRELIMINARY: PRT: 449 AA.

AC 022822: 01-JAN-1998 (TEMBLrel. 05, Created)

us-09-147-955-6.rsp

DR	PROSITE	PS00375; UDEPT; 1.
DR	PFAM; PF00201; UDEPT; 1.	
	Transferrase.	
SO	SEQUENCE	479 AA; 53913 MW; 2ED16CCC CRC32;
Query Match	29.6%	Score 711.5; DB 10; Length 479;
Best Local Similarity	35.8%; Pred. NO. 4.1e-47;	

```

Query Match      29.6%   Score 711.5; DB 10; Length 475;
Best Local Similarity 35.8%   Pred. 10.4e-47;
Matches 172; Conservative 89; Mismatches 174; Indels 45; Gaps 15;

QY    5 HLLATPPAGCHINPLAIOFAKRLANADQVPEFNSVYAMRMSSTLAGSN-----G 55
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::  |
Db     8 HHMLVSPDGGGHNPILRLGLKLASKGLAVTFVTTERKKMKKKHQAQKIDGVLRKYGAS 67

```

```

QY 56 LINFVSFGYDDGLOPGDDGKKNYSEMSKRGIKALSDTLAANNVDGSSK-TTFVYSH 114
Db 68 FIREPFSOGFAD-----DDEKRPFDEAFRPHLEAVGKEIKNLVKRYNEKPEYTCILNNA 122
QY 115 LFMAAAVAEAEFLRBSALLMIERPAVLIDFPYFNGY-----SDEIDAGDAIHLEGGP 169
Db 123 FVPVACVAEELHPSAIVLVWASCACTAAYVYIHHRLKFEPTKEPDI---SVEIP-CP 178
QY 170 VLAORDPFSFLPST-HERFRSLIMEKETLEGECKPKVPLVNSFDALPEPALKAIDRYEM 228
Db 179 LKRADEISFLPHSSPYTAFGDIIDDLKRFENHKSFLYFDIFPRELEKIDIMHMSQLCP 238
QY 229 IAI-----GLIPSAFLFDGKDPDRSGDLEFKGNDDDCLEMISTPSSSVYVYSGFSF 284
Db 239 QALISPYGPEFLPKMAQTLSSDVK-----GDISEPAS--DCKEMIDSKSPSSVYISFGTI 290
QY 285 VNTTKSOMEIARGLDLCGRPELWVR-VNEGEVYLISC-EELKRGKIVSMCSOLEV 342
Db 291 ANLKOQEMEIARGLVSLSGSLVLMVVRPMECTFVEPVHLPRELEEGKIVEMCPQERVL 350
QY 343 THPSLGFVHCCGMSNTLISISGCVMAVFPMPGCGNATLMDKVRCTVRV-RANEG 401
Db 351 AHPAIPACFLSHCCGMSNTLHTRGVAVVCFPPMGQVYDVAIDLAVRTVGRGKAABE 410
QY 402 SVYDGEIDIRICIEYVDG--GEKSRRLRESAGKMWDLARKAMEEDGSSVNNLAVFLDEVY 459
Db 411 MIVSREVY--AEKLTLEATYGERAVELFENARMRKAEAEAAVADGSSDMNFKFEYDKLV 467

RESULT 14
OZ2182 PRELIMINARY; PRT; 456 AA.
ID 022182
AC 022182;
DT 01-JAN-1998 (TEMBREL. 05, Created)
DT 01-JAN-1998 (TEMBREL. 05, last sequence update)
DT 01-NOV-1999 (TEMBREL. 12, last annotation update)
DE PUTATIVE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
GN T20D16.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROBINSEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA STYES S.M., KAUI S., MASON T.M., KERLAUAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002391; AAB87119.1; -
DR MENDEL; 26261; Atrah; j145;26261.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 456 AA; 50713 MW; 9A33B1F9 CRC32;

```

Best Local Similarity 36.1%, Pred. No. 5, 1e-46;

Matches 167; Conservative 93; Mismatches 175; Indels 28; Gaps 14;

```

OY 5 HVLLATFPAGIINPALOPAKR--LANADIVTFTSVYAMRMSRTAAGSNGLIINVSF 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 HVLMTVLPFGGIIINPLAKLKLSSKNHINLAT--IEARDLSTVEKRPVPLVVF 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 63 SDGYDGLDOPGDGKNVSEMKSRGIRKALSDTLAANNVDO--KSKITFVYSHLEPAMA 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 69 SDGL-----PKEDPKAPETLKLK-----LAKVGMMLSKIEERKSCIISSPFPVWP 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 121 KVARREHLMSALLMTBPATVLDIEYTF--NGYSDEIDAGSDAHLPGSLPVLQORDP 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 AVASNNICALIMIQAGAVSVYRYMKTNSPFDLEDL--NQVLELP-ALPILLEVRD 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 178 SFPLSTHFRPSLAKKELETLEGEEKPVLVNSFDALPDALKAI--DKYEMIAIGPLP 236
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 SFPLSPGGAHFTNLMAEFADCLARYK--WVLYNSFYLESEIIIESMADLKVPVPIPLV- 232
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 237 SAEPLDGRPSDSFEGDLPFEKGSNDDCLEWLSTNPRSSVYVVSFGSFVNTTKSOMEETA 296
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 SPFLTG--DDEETLDGKNLDFCKSDDCMELDKQARSVYIISFGSMLETTLENOYETIA 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 297 RGLDGRGPEFLWVAVRNEGESEYVLSCEMELKR--VGKIVMSOLEVLTSPSLGCFTHG 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 292 KALKNGRLPEFLVIRPREKAQVAVLOEMVKEGCVLEMSPOEKILSHAIICFTVTHG 351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 356 WNSTLESIFGVPAVAFPOWEDGNTAKLMEDVWRTGVRVANEBSVVDGDIRRCIEE 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 352 WNSTMETVAVGVVAVPSWTDPIDARLVDVFGVGRNRKNSVDELKVEEVERCIEA 411
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 416 VMDGKSRKLRRESAGKMDLARKAMEEDGSSVNNLKYFLDEV 458
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 412 VTE--GPAVDIRRAELKRVARLALAPGSSSTRNDLFTSDI 453
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 15

```

ID 022183 PRELIMINARY: PRT: 438 AA.
AC 022183;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 12, Last annotation update)
DE PUTATIVE INDOLF-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
GN 120D16.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA ROUSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAJAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002391; AAB87106.1; -
DR MENDEL: 26262; Arab73145; 26262.
DR PROSITE: PS00375; UDPGT; 1.
DR PFM: PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 438 AA: 48717 MW: 88F76990 CRC32;

```

Query Match 28.3%, Score 682; DB 10; Length 438;

Best Local Similarity 36.2%, Pred. No. 6, 9e-45;

Matches 166; Conservative 81; Mismatches 168; Indels 44; Gaps 16;

```

OY 14 OGHINPALQFAARLANADIVTFTSVYAMRMSRTAAGSNGLIINVSFSDGYDGLDLP 73
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 7 OGHINPLMKFAARLANADIVTFTSVYAMRMSRTAAGSNGLIINVSFSDGYDGLDLP 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

OY 74 DQGN---YSEKSRGIRKALSDTLAANNVQKSKITFVYSHLEPMAKVARREHLMS 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 DDDRDPTLAKSLKDGAKNLKIEERFD-----CLISVFPFWPVAVAANNIRP 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 131 ALTIMPAAVLDIEYTF--NGYSDEIDAGSDAHLPGSLPVLQORDLPSPFLSTHER 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 115 AILMIQAGAFSVYRYMKTNPFDLEDL--NQVLELP-ALPILLEVRDPSLMLPGGAN 172
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 188 FRLMKKELETLEGEEKPVLVNSFDALPDALKAI--DKYEMIAIGLISAFLDGKDS 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 173 VNTLMAEFADCL--KDKVVLVNSFYLESEIIIESMDLKPIPIGLV--SPFLGNDDE 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 247 DRSGDLEFEKGSNDDCLEWLSTNPRSSVYVVSFGSFVNTTKSOMEIARGLDGRPF 306
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 -----KTLDMKVVDDCYCEMLDKQARSVYIISFGSLKLENOYETIATALKNGVFP 283
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 307 LMYRYANE--GEYVLSCEMELRVCK--IVSMCSOLEVLTSPSLGCFVTHGNNSTLES 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 LMYIRREKGENYQV--IQEMVKEGCVVTEMGQOEKILSHAIISCFITCGNNSTIEY 341
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 364 SFGVPVAFPOWEDGNTAKLMEDVWRTGVRVANEBSVVDGCD---EIRRCIEEYMDG 419
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 VTGVPVAVPTWIDQPLDARLVDVFGVGRNK-----NDALDGLKVAEVERCIEAVTE- 396
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 420 GERSKRLRESAGKMDLARKAMEEDGSSVNNLKYFLDEV 458
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 397 GPAADMRRAATELKHARSAMSPGSSAQNLDFTSDI 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: August 1, 2000, 21:31:36
Job time: 6065 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:21 ; Search time 47.63 Seconds

(without alignments)
229.252 Million cell updates/sec

Title: US-09-147-955-6

Perfect score: 2407

Sequence: 1 MSRAHVLLATFPAGSHINPA.....EEDGSSVNNLKVFLEDEVGCI 461

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2407	100.0	461	1	W92949
2	1646.5	68.4	460	1	W92947
3	1583.5	65.8	443	1	W92948
4	1473.5	61.2	478	1	W92950
5	1320	54.8	468	1	W92952
6	773.5	32.1	466	1	W09825
7	751.5	31.2	471	1	W03756
8	431	17.9	470	1	W47172
9	429.5	17.8	433	1	W13406
10	425	17.7	471	1	W98009
11	405.5	16.8	488	1	W64569
12	377	15.7	473	1	W56451
13	340.5	14.1	471	1	R49245
14	147	6.1	515	1	W56750
15	140.5	5.8	506	1	R10429
16	136.5	5.7	414	1	R93982
17	135	5.6	330	1	W47126
18	135	5.6	330	1	R26153
19	125	5.2	534	1	R26154
20	103	4.3	74	1	R30165
21	102	4.2	473	1	W64392
22	102	4.2	1198	1	W64384
23	94.5	3.9	209	1	R85453
24	94	3.9	431	1	R07464
25	94	3.9	431	1	W87890
26	92	3.8	1974	1	W98391
27	91.5	3.8	474	1	R88093
28	90.5	3.8	434	1	R13493
29	90.5	3.8	479	1	R19442
30	88.5	3.7	413	1	W93141
31	88	3.7	365	1	R28827
32	86.5	3.6	460	1	W36114
33	86.5	3.6	800	1	W34479

34	85.5	3.6	1093	1	R42818	TMF. New protein c
35	84	3.5	816	1	R85870	WD-40 domain-contg
36	83	3.4	492	1	W34994	Archaeobacterium AE
37	83	3.4	804	1	W22709	Leucyl-tRNA synth
38	83	3.4	957	1	W20841	H. pylori transmem
39	83	3.4	1766	1	W24790	P. falciparum live
40	82.5	3.4	408	1	W71464	Glycosyltransferas
41	82.5	3.4	578	1	W10422	Penicillium chrys
42	82	3.4	417	1	W54358	Phosphoglycerate k
43	81.5	3.4	563	1	W56476	Amino acid sequenc
44	81.5	3.4	855	1	W01601	Aquifex pyrophilus
45	81	3.4	444	1	W90340	G. max truncated S

ALIGNMENTS

RESULT 1
ID W92949 standard; Protein; 461 AA.
AC W92949:
DE W09905287 Seq ID 9.
KW Plant; Flavonoid 5-transglycosylation activity; 5tg; variety; colour.
OS Verbena hybrida.
PN W09905287-A1.
PD 04-FEB-1999.
PE 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR.) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR N-PDB; X02828.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Claim 2; page 76-78; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5tg) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC flowers harvested from them.
SQ Sequence 461 AA;

Query Match 100.0%; Score 2407; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 3; 6e-250;
Matches 461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSRAHVLLATFPAGSHINPALQFAKRLANADIVTFTSYAMRRMSRTAAGNGLINFEV	60
DB	1	MSRAHVLLATFPAGSHINPALQFAKRLANADIVTFTSYAMRRMSRTAAGNGLINFEV	60
QY	61	SSSDYDGDGDPGDDGKMYSEMSKRGIKALSDPLAANNVDKSSITFYVYSHLPANAA	120
DB	61	SSSDYDGDGDPGDDGKMYSEMSKRGIKALSDPLAANNVDKSSITFYVYSHLPANAA	120
QY	121	KVAREFHRLSALWIEPAVVDIFYFYFNGYSDEIDAGSDAHLPGGLFVLAORDLPSEFL	180
DB	121	KVAREFHRLSALWIEPAVVDIFYFYFNGYSDEIDAGSDAHLPGGLFVLAORDLPSEFL	180
QY	181	LPSTHERFRSLMKETLEEGEEKPYLVNSPALPEMDAKIDKEMALICPLIPSAFL	240
DB	181	LPSTHERFRSLMKETLEEGEEKPYLVNSPALPEMDAKIDKEMALICPLIPSAFL	240
QY	241	DKKDDSDRSFGGDLFEKGSNDDCLEWLSTNPRSSVYVYFSGFVWTTKSMOEIARGL	300
DB	241	DKKDDSDRSFGGDLFEKGSNDDCLEWLSTNPRSSVYVYFSGFVWTTKSMOEIARGL	300
QY	301	DGGRFLWVVRVNGEVELISCMELKRVGKYSKQLEVLTHPSLIGCFYHCGWNSITL	360
DB	301	DGGRFLWVVRVNGEVELISCMELKRVGKYSKQLEVLTHPSLIGCFYHCGWNSITL	360

OY 361 ESTSPGVAVAPPOWEDOGTNAKIMEDVYRTGVVRANEGSVVDDETRCIEEVMOG 420
 DB 361 ESTSPGVAVAPPOWEDOGTNAKIMEDVYRTGVVRANEGSVVDDETRCIEEVMOG 420
 OY 421 EKSRLRESAGKMKDLARKAMEEDGSSVNNLKFLDEVG1 461
 DB 421 EKSRLRESAGKMKDLARKAMEEDGSSVNNLKFLDEVG1 461

RESULT 2

W92947
 ID W92947 standard; Protein; 460 AA.
 AC W92947;
 DT 14-MAY-1999 (first entry)
 DE MO905287 Seq ID 7.
 KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
 OS Petilla frutescens.
 PN MO905287-A1.
 PD 04-FEB-1998.
 PF 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNK) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 Yamazaki M.
 PI WPI: 99-142940/12.
 DR N-PSDB: X02826.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 used to transform plants for improvement of plant coloration
 PS Claim 2; Page 71-73; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have
 flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
 in the production of plant varieties with improved colour, for example in
 CC the production of flower varieties with desired coloration and cut
 CC flowers harvested from them.
 SQ Sequence 460 AA;

Query Match 68.4%; Score 1646.5; DB 1; Length 460;
 Best local Similarity 70.0%; Pred. No. 4.4e-168;
 Matches 326; Conservative 49; Mismatches 74; Indels 17; Gaps 7;

OY 1 MSRAHYLLATFPAGHINPALOFARLANADIOVFTFTSYANRRSR---AAGNGLI 57
 DB 1 MRRRLVLTFFPAGHINPALOFARLANADIOVFTFTSYANRRSR---AAGNGPGL 60
 OY 58 NFVSFSDGYDDGLQPGDDGKNYSEMKSRGKALSDTLAANNVDKSSKTFVYSHLFA 117
 DB 61 DFVAFSDGYDDGLQPGDDGKNYSEMKSRGKALSDTLAANNVDKSSKTFVYSHLFA 114
 OY 118 MAKVAREEHLASALLMEPATVLDIFEFYNGYSDIDGSDAHLPGSLPLAQRDL 177
 DB 115 MAEVARRESQVSAALLMEPATVLCITYFNGYADIDGSDIQLP RLPLEGRSLP 173
 OY 178 SFLLPSTHERFRSLKKEKLETLLEGEEKVLYNSFDALPDAKAIKDKEMIAIGPLPS 237
 DB 174 TELLPTPERFLMKKEKLETLLEGEEKVLYNSFDALPDAKAIKDKEMIAIGPLPS 233
 OY 238 AFLDGPDSRSGDLEKSGNDDEMLSTNPRSSVYVSGSVNTTSQMEIAR 297
 DB 234 AFLDGPDSRSGDLEKSGNDDEMLSTNPRSSVYVSGSVNTTSQMEIAR 292
 OY 298 GLDLCGRPLLVYR-----NDEGEVLISCMELKRVKIVSMCSOLEVLTHPSLGF 352
 DB 293 GLDLCGRPLLVYR-----NDEGEVLISCMELKRVKIVSMCSOLEVLTHPSLGF 352
 OY 353 HCGMNSTLESISFGVPMAPFOWDOGTNAKIMEDVYRTGVVRANEGSVVDDETRC 412
 DB 353 HCGMNSTLESISFGVPMAPFOWDOGTNAKIMEDVYRTGVVRANEGSVVDDETRC 411
 OY 413 IEENVNDGERSKRLRESAGKMKDLARKAMEEDGSSVNNLKFLDEVG 458
 DB 412 IEENVNDGERSKRLRESAGKMKDLARKAMEEDGSSVNNLKFLDEVG 457

RESULT 3

W92948
 ID W92948 standard; Protein; 443 AA.
 AC W92948;
 DT 14-MAY-1999 (first entry)
 DE MO905287 Seq ID 8.
 KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
 OS Petilla frutescens.
 PN MO905287-A1.
 PD 04-FEB-1998.
 PF 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNK) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 Yamazaki M.
 PI WPI: 99-142940/12.
 DR N-PSDB: X02827.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 used to transform plants for improvement of plant coloration
 PS Claim 2; Page 74-76; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have
 flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
 in the production of plant varieties with improved colour, for example in
 CC the production of flower varieties with desired coloration and cut
 CC flowers harvested from them.
 SQ Sequence 443 AA;

Query Match 65.8%; Score 1583.5; DB 1; Length 443;
 Best local Similarity 68.7%; Pred. No. 2.4e-161;
 Matches 310; Conservative 57; Mismatches 65; Indels 19; Gaps 7;

OY 1 MSRAHYLLATFPAGHINPALOFARLANADIOVFTFTSYANRRSR---AAGNGLI 57
 DB 1 MRRRLVLTFFPAGHINPALOFARLANADIOVFTFTSYANRRSR---AAGNGPGL 60
 OY 58 NFVSFSDGYDDGLQPGDDGKNYSEMKSRGKALSDTLAANNVDKSSKTFVYSHLFA 117
 DB 61 DFVAFSDGYDDGLQPGDDGKNYSEMKSRGKALSDTLAANNVDKSSKTFVYSHLFA 114
 OY 118 MAKVAREEHLASALLMEPATVLDIFEFYNGYSDIDGSDAHLPGSLPLAQRDL 177
 DB 115 MAEVARRESQVSAALLMEPATVLCITYFNGYADIDGSDIQLP RLPLEGRSLP 173
 OY 178 SFLLPSTHERFRSLKKEKLETLLEGEEKVLYNSFDALPDAKAIKDKEMIAIGPLPS 237
 DB 174 TELLPTPERFLMKKEKLETLLEGEEKVLYNSFDALPDAKAIKDKEMIAIGPLPS 233
 OY 238 AFLDGPDSRSGDLEKSGNDDEMLSTNPRSSVYVSGSVNTTSQMEIAR 297
 DB 234 AFLDGPDSRSGDLEKSGNDDEMLSTNPRSSVYVSGSVNTTSQMEIAR 292
 OY 298 GLDLCGRPLLVYR-----NDEGEVLISCMELKRVKIVSMCSOLEVLTHPSLGF 350
 DB 293 GLDLCGRPLLVYR-----NDEGEVLISCMELKRVKIVSMCSOLEVLTHPSLGF 352
 OY 351 YHCGMNSTLESISFGVPMAPFOWDOGTNAKIMEDVYRTGVVRANEGSVVDDETR 410
 DB 353 YHCGMNSTLESISFGVPMAPFOWDOGTNAKIMEDVYRTGVVRANEGSVVDDETR 411
 OY 411 RCIEENVNDGERSKRLRESAGKMKDLARKAM 441
 DB 412 RCIEENVNDGERSKRLRESAGKMKDLARKAM 442

RESULT 4

W92950
 ID W92950 standard; Protein; 478 AA.
 AC W92950;
 DT 14-MAY-1999 (first entry)
 DE MO905287 Seq ID 10.
 KM Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.


```

PE 16-JUL-1998: J03199.
PR 25-JUL-1997: JP-200571.
PA (SONR.) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
PI WPI: 99-142940/12.
DR N-PSDB: X02831.
PT Gene encoding a plant flavonoid 5-transglycosylase - can be
PT used to transform plants for improvement of plant coloration
PS Claim 2; Page 83-85; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylase (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SO Sequence 468 AA:

Query Match 54.8%; Score 1320; DB 1; Length 468;
Best Local Similarity 55.7%; Pred. No. 5,2e-133;
Matches 264; Conservative 76; Mismatches 108; Indels 26; Gaps 10;

OY 1 MSRAHVLAFPPAGGHINPALOPAKRLADIAOVFFTSYVAMRMRSTA--GSKGLIN 58
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MVQGHVILTFPPAGGHINPALOPAKRLAKYMGIEVFTSISIAQSRMDEKILNAPGL-N 59

OY 59 FVSRSDGIDYDGLPGDDGKNINSEKSRICAKLSDTLANNVDKSKITFVYSHLPFW 118
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 60 FIPSPDGFDEGFHSKDPVFMYSOLRKCGSETVKKIIL--TCSENGOPITCLLYSIFLPW 117

OY 119 AAKVAREFHSLRALIEPATVLDIFYEYFNGY----SDEIDACSDAHLPGGLPYLAOR 174
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 118 AAEEVAREVHILPSALLMSQPATILDIYFNHGYEKAMNENSDNMSIQLP-GLPILETR 176

OY 175 DLPSFLLPSTHEFRSL-----MKKLETLEGEEKPVYLVNSDALPALAKAIKYM 228
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 177 DLPSFLLP--YGAKGSIKRVAPLPPFKELIDTLAETTPILVNTFDELEPALNAIGFAR 234

OY 229 IATGPIPLPSAFLGKPPSDRSFGGDLFEKGSNDDCLEMLSTNRSVVSVVSGFSFVNT 288
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 235 YGIGPIPLPSAFLGNDPLDASFGGDLFQ--NSNDYEMMLSKNSKNSSVYISGLMANS 291

OY 289 KSQMEELARLLDCCGRPILMYVRV---EGEVLISCMELKRVKGVIVWCQGLEVLN 344
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 292 ISQMEELSKLLIDIGRPILMYIKENKEGKEENKKLGIELEIKIGIVWCQGLEVLN 351

OY 345 PSLGCFEYHCGWMSSTLESISFGVPVAPFPOMDQGTNAKLMIEDVWRGVVARNBEGSYV 404
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 352 PSLGCFEYHCGWMSALESLACGVPVAPFPOMDQGTNAKQVEDWVKGSQVAVRLINEDG-VY 410

OY 405 DGDDIRRCIEVDVNDGGEKSRKLRESGKWKDLAKARAMEGSSVNNIKVYLDEY 458
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 411 ESEIEIKCIELVMDGEGKEBELKKNAKKWKELAREAVYEGSSSHRNKLAVIDDY 464

RESULT 6
W09825
ID W09825 standard; Protein: 466 AA.
AC W09825;
DC 15-JUL-1997 (first entry)
DE UDP-glucose:thiohydroximate S-glucosyltransferase.
KW UDP-glucosylate; UDP-glucose:thiohydroximate S-glucosyltransferase;
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola.
OS Brassica napus cv. Westar.
FH Key Location/Qualifiers
FT misc-difference 2
FT /note="residue 2 is Val in other S-GT isoforms"
FT /note="a Lys residue is inserted between amino
FT /note="acids 10 and 11 in some S-GT isoforms"
FT misc-difference 12
FT /note="residue 43 is Ser in some S-GT isoforms"
FT misc-difference 43

```


RESULT	8
ID	W47172 standard; protein: 470 AA.
AC	W47172
DE	08-JUN-1998 (first entry)
DE	Glucosyl transferase (GTase) protein encoded by TW11 gene.
KW	Glucosyl transferase; GTase; TW11 tomato; signaling pathway;
KW	salicylic acid; jasmonic acid; ethylene; wound inducible gene;
KW	plant defence protein; plant response; tobacco; rice.
OS	Lycopersicon sp.
PN	M09745546-A1.
PD	04-DEC-1997.
PE	30-MAY-1997.
PR	31-MAY-1996; GB-011420.
PA	(UYUO-) UNIV YORK.
PI	Bowles DJ, Calvert CM, O'Donnell PJ, Roberts MR;
DR	WPI: 98-032653/03.
DT	WPI: V17054.
RT	Tomato wound inducible (TW11) gene encoding glucosyl transferase

```

0Y 1 MSRAHVILATFPAGOHINALOPARLANADIOVTEFT-----SYAAMRMRKTAASNG 55
Db 1 MGEIHFEFFPDQOHMIPITLDMANVAVRGVKTITITPLNEBSVSAKIERKHJGIEI 60
0Y 56 LINFVSDEYD-----GLQGDG--KNVASEKSRGICALSDTLAANNVDOKSG 106
Db 61 DILRKPAENDELPEDCRDLVPSODKLPNFLK-----AAAMKREFSEL 107
0Y 107 I-----TEVVYSHLEFMAAKVAREHLSALLMEPTVDIDITYEFNGSDIDAGSDA 161
Db 108 IGECRPDCLVSDHFLPMTYDSAKFSI-----PRIV-----FHQSYFALCVDT 152
0Y 162 IHLGGGPIVLAQDPSFLPSTHERFRLKKELEFLEGEERK-----206
Db 153 IRRKKPKKNSS-DITETVVPDLPHETR-LRLODSFEBSDEDETGMAMIAVRESAK 211
0Y 207 ---VLVNSFDALPPDALKAIDKYEHI-----AIGLIPSAFLDGKDPDSFGDLEK 255
Db 211 SYGIEFMSFELESD---VEHYHTVVGRRKMAAGPL---SLCNRDIEDKNERG---RK 260
0Y 258 GSNDD-CEMLISTNRSSVVYVSGSPVNTKCSOMEIARGLLDGRFLVWVNVNGE 310
Db 261 SSIDENHACKLMDOKSSSIYVCGSTADTTQOMBELAMGLBAGOPFVITGND 320
0Y 317 EVLISCHEELKRVGKIV-SWCSQLEVYLTHPSLGCFTHGCWNSTLESISFGVPWATPQM 377
Db 321 WLPGEBERTKENGILLINMAPOSYLIDHEHIGAFTVHCMSNNTLGISAGVPMTWYF 380
0Y 376 FDOGTNKLMEVDWRTGVAV-----RANEGBVYDDELTRRICEVMDGCKSRKLES 429
Db 381 AEOFPNKKVLTVEMRGAOVSQKMKRATASG--YKREALAKIKRYM-ASETJGFSR 433
0Y 430 AGKKWDLARAMEEDGSSVNNLKYFLDEV 458
Db 438 AKYEKEMAREALIEBGGSSYNGWATLIDOI 466

```

RESULT	9
W13406	
ID	W13406 standard; Protein; 433 AA.
AC	W13406; 1997
DT	10-JUN-1997 (first entry)
DE	Solanum melongena flavanoid-3-glucosyl transferase.
KW	Esopplant; flavanoid-3-glucosyl; transferase; glucosae; flavonoid;
OS	Pigment; production; seed; plant.
PN	Solanum melongena.
PD	J09056385-A.
ED	04-MAR-1997.
PF	25-AUG-1995; 238943.
PR	25-AUG-1995; JP-238943.

PA (KIRI) KIRIN BREWERY KK.
DR WPI: 97-206631/19.
DR N-PSDB: 162123.
PT Solanum melongena flavonoid-3-glucosyl transferase and DNA encoding
PT 1t - catalyses transfer of glucose to 3-position of flavonoid
PT pigment in plants
PS Claim 4; Pages 9-10; 14pp: Japanese.
CC The present sequence is the Solanum melongena (eggplant)
CC flavonoid-3-glucosyl transferase (F3GT), which catalyses the
CC transfer of glucose to the 3-position of flavonoid pigments. The
CC F3GT DNA is useful for the production of the seed of a plant. A
CC 5. melongena seed was cultured under a red light and a flavonoid
CC derived. A cDNA library in which the pigment was concentrated was
CC prepared. A flavonoid glucosidizing enzyme gene was cloned and its
CC structure and mRNA expression pattern analysed. The clone was
CC expressed in E. coli.
SO Sequence 433 AA;

Query Match 17.8%; Score 429.5; DB 1; Length 433;
Best Local Similarity 26.7%; Pred. No. 1.9e-37;
Matches 123; Conservative 89; Mismatches 189; Indels 59; Gaps 15;

QY 2 SRAHYLATPFAOGHINPALOFARLANADIQVTFPTSYVAMRMSRTAGSNGILNFS 61
DB 4 SOLHATLAFPFPGTHATPLTLVOKISPLPSTIFS-----FNITSSNSSTFSKVP 56
QY 62 PSDG---YD--DGLQPDGDKNYSEKSKGICALSDTLAANNVDKSKITFV---VY 112
DB 57 NOENIKIYVWDYEGCNDPFFGLEAIR---LFIQSTLISKITTEAEELGVFSCIF 112
QY 113 SHLFAMA--AKVAREPHLSALLMTEPATVLDIFYFYNGSYDEIDAGSD-AIHPGGLP 169
DB 113 SDAFLMCLVLVLPKKNMARGVAYWTGGSSCLAVHL-----YDILINSKNETSLKIPGSS 167
QY 170 VLAQDLPSEFLPSTHERFSLMEKEL--ETLGEERPKVLYNSFALPEPAL--KAIDK 225
DB 166 TLISINDIPPEV---TAELLEGPMSSMLYNMNLHRAADAVVLSFGLDRDLINKDKQ 224
QY 226 --YEMIAIGPLIPSAFLDKDPSDRSGDGLFEKGSNDDEMLESTNPRSSVYVVSFGS 283
DB 225 NQKVFNIQPLVLOS-----SRKIDESCIOIMDKQEKSVVLSLGT 267
QY 284 FVNTKRSQMEELAKGLDCCGRPLVAVRVNEGEVULISCEBELKRYGKIVSCOLEVLT 343
DB 268 VTTLPNIGSIAELLETKKPFTMSLRNNGVKNLPKGLERTKEFGKIVSWAPOLIELA 327
QY 344 HPSLCGFYTHGGMNSTLESISFGVPVAFQPMFQDGTNAKLMEDVARTGVRRANDEGSV 403
DB 328 HKSVCVFYTHGGMNSTLESISFGVPNCRPFQDQKMSWESVWEILOI----EGGI 383
QY 404 VGDRLRCIEVDGDKSKRLKESAGKKDLAKKAMEE 443
DB 384 FTKSGIISALDTFFN-BEKKILRENEGLKALEKALEVANO 422

RESULT 10
W98009
ID W98009 standard; Protein; 471 AA.
AC W98009;
DT 21-JUN-1999 (first entry)
DE Medium-chain UDP-glucose:fatty acid glucosyltransferase.
KW UDP-glucose:fatty acid glucosyltransferase; tomato;
KM 1-O-acyl-a-glucose; acyl donor; esterification.
OS Lycopersicon pennellii1.
OS Lycopersicon esculentum.
PN W09090144-A1.
PD 23-FEB-1999.
PF 10-AUG-1998; 1B1369.
PR 30-JUN-1998; US-106464.
PR 13-AUG-1997; US-055554.
PA (CORR) CORNELL RES FOUND INC.
PI Bananella NT, Ganges GS, Kull J, Steffens JC;

DR WPI: 99-190155/16.
DR N-PSDB: X24873.
PT New chain length specific uridine diphosphate-glucose:fatty acid
PT glucosyltransferases - useful for preparing straight or branched,
PT (un)saturated 1-O-acyl-a-glucoses and 1-O-acyl-a-D-glucoses
PS Example 3; Page 36-38; 42pp: English.
CC The present sequence is a UDP-glucose:fatty acid glucosyltransferase
CC that transfers glucose to medium-chain length fatty acids to form
CC 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses act as acyl donors
CC in the esterification of glucose and in the further esterification
CC of partially acylated glucose in reactions catalysed by glucose
CC acyltransferases. They also act as acyl donors in the
CC esterification of sucrose and in the further esterification of
CC partially acylated sucrose in reactions catalysed by sucrose
CC acyltransferases. A cDNA clone (see X24873) encoding the enzyme
CC was isolated by immunoscreening a leaf trichome cDNA library
CC prepared from an F1 population of a cross between wild tomato
CC Lycopersicon pennellii (LA 716) and cultivated tomato
CC Lycopersicon esculentum cv. New Yorker. The invention provides
CC chain length-specific UDP-glucose:fatty acid glucosyltransferases
CC and a method for their purification. The enzymes have specificity
CC for transferring glucose to short, medium or long chain length
CC fatty acids. Methods for preparing 1-O-acyl-a-D-glucoses, where
CC the acyl group comprises 3-5, 6-13 or 14-22 C atoms, are claimed.
SO Sequence 471 AA;

Query Match 17.7%; Score 425; DB 1; Length 471;
Best Local Similarity 27.9%; Pred. No. 6.6e-37;
Matches 144; Conservative 71; Mismatches 193; Indels 108; Gaps 19;

QY 1 MSRAHYLATPFAOGHINPALOFARLANADIQVTFPT---SYVAMRMSRTAGSNG 55
DB 2 MQLHFFPPMAQGMHPTLDMAKIVACRGVAKITTTPLANSYSKAIERKKHIGIET 61
QY 56 LNFVFSFSDGYD-----GLQPDGDK--KNYSEKSKGICALSDTLAANNVDKSK 106
DB 62 DIRLKFPAKENDLPEDERLDLPVSDDKLPNFK-----AAAMWKDEFEL 108
QY 107 I-----TFVYSHLPAMAKVAREPHLSALLMTEPATVLDIFYFYNGSYDEIDAGSDA 161
DB 109 IGCRCDCVLSMFLPMTDSNAKFSI-----PRVYH-----GTGY 145
QY 162 IHPGGLPVLAOR-----DLPSEFLPSTHERFSLMEKLETLGEERPK----- 206
DB 146 FALCVGHSIRKRPKNVSDTEFVVPDPPEHTR-LTRQLSPFQSDSEEGMAPMIKA 204
QY 207 -----YVNSFDLPEPALKAIDKYEML-----AIGPLISAFIDKDPDRSF 250
DB 205 VNESDAKSGVIFNFFLEED---YVEHTYKVGKRNMAIGPL-----SLCNRDIEYKAE 257
QY 251 GGDLEFKGSNDG--CLEMLSTNPNSSVYVVSFGSEVNTTKSQMEELARGLLDCCGRPLVAV 309
DB 258 RG---RKSSIDHACLKWLDSKSSIVYVCFGSTADFTTAQMOELAMGLEASGODFIWY 314
QY 310 VRYNEGEVULISCEBELKRYGKIV-SMCSOLEVLTTPHSLCFTYHGMNSTLESISFGVP 368
DB 315 IRTGNEDMLPEGEERTKEKGLIIRGNAPQVLLIDHAIATPFTYHGMNSTLEGISAGVP 374
QY 375 MTLVPYFAQFNEKTLVTEVRSAGVSGSKOMKRTASEG--VRREAIARIRVM-ASPE 431
QY 423 SRRLRESAGKMKDLARKAMEDESSVNNLKVFLOEY 458
DB 432 TEGFNRKAEYKEMAREALIEGGSSYNGMATTIIDI 467

RESULT 11
W64569
ID W64569 standard; Protein; 488 AA.
AC W64569;
DT 07-DEC-1998 (first entry)

DE	Potato	solanidine UDP-glucose glucosyltransferase.
KM	Solanidine UDP-glucose glucosyltransferase; SCT; potato; tomato; glycoalkaloid; solasodine; solanaceous plant; transgenic plant.	
OS	Solanum tuberosum cv. Lemni Russet.	
PH	Key	
FT	Region	
FT	12..46	
FT	/note="conserved domain"	
FT	110..143	
FT	Region	
FT	/note="putative steroid binding region"	
FT	351..401	
FT	/note="UDP-glucose binding region"	
PN	WC9834471-A1.	
PD	13-AUG-1998.	
PE	30-JAN-1998: U01864.	
PR	07-FEB-1997: US-797226.	
PA	(USDA) US SRC OF AGRIC.	
PI	Allen P.V, Belknap W, Friedman M, Moehs CP, Rockhold DR.	
PI	Stapleton A;	
DR	WPI: 98-446828/38.	
DR	N-PSDB: V49609.	
PT	New DNA encoding solanidine UDP-glucose glucosyl-transferase and related vectors and transgenic plants, used to reduce contents of steroidal glyco-alkaloids) in plants, specifically potato	
PT	steroidal glyco-alkaloids) in plants, specifically potato	
PS	Example 1; Fig 2; 54pp; English.	
CC	This is the amino acid sequence of potato cv. Lemni Russet	
CC	solanidine UDP-glucose glucosyltransferase (SCT). It was deduced from an SCT cDNA sequence (see V49609). SCT is involved in the biosynthesis of steroidal glycoalkaloids in solanaceous plants, catalysing the UDP-glucose dependent conversion of the aglycone solanidine to gamma-chaconine. The invention relates to DNA sequences which encode SCT, and their use, particularly use of an antisense DNA construct to inhibit SCT activity and glycoalkaloid levels in solanaceous plants. Transgenic plants are claimed, particularly tomato and potato, that have reduced contents of toxic steroidal glycoalkaloids. The inactivation of glycoalkaloid biosynthetic pathways is beneficial to reduce or eliminate glycoalkaloid biosynthesis during storage and shipping.	
CC	Sequence 488 Aa;	

Db 405 GCGTCGIDGYNMEGEI-----TGCVISAKIKREAIERIMISNGSEETITINRVMAMSKM 460
QY 437 ARKAMEEGGSSVNNIKVFLDEY 458
| : | | | | : :
| : | | | | : :
Db 461 AQNATNEGSSWNNTALIOHI 482

RESULT 12
W56451
ID W56451 standard; Protein; 473 AA.
AC W56451;
DT 06-AUG-1998 (first entry)
DE UDP-glucose:flavonoid 3,5-O-glucosyl transferase amino acid sequence.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
KW blue flower.
OS unidentified.
PN J1013184-A.
PF 06-MAY-1998.
PR 15-OCT-1996; JP-272253.
PA (TOLG) TOKYO GAS CO LTD.
DR WPI; 98-315473/28.
FA
PI PSDB; V23108.
PT Gentian flavonoid glucosyl transferase gene - used to produce
PT recombinant rose plants with blue petal(s)
PS Claim 1; Pages 6-7; 10pp; Japanese.
CC The present sequence represents a udp-glucose:flavonoid 3,5-O-glucosyl/
CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (slc) petals. The gene is
CC used to transform rose plants to produce blue flowers.
SQ sequence 473 AA;

Query Match	15.78;	Score 377;	DB 1;	Length 473;
Best Local Similarity	24.88;	Pred. No. 9.5e-32;		
Matches 130; Conservative	90;	Mismatches 184;	Indels 120;	Gaps 17,

Query Match	16.8%	Score 405.5	DB 1	Length 488
Best Local Similarity	26.1%	Pred. No. 8.7e-35		
Matches 131	Conservative	93	Mismatches 199	Indels 79
				Gaps 18
QY	5	HLVLAPEPAGGHINPALQFAKRIANADIDVTFEFTVY--AMRMSTRAAGSNGL-----	56	
Db	12	HLVLEPLPLASGHFFPLVYNARLFJASNGVAKATILITPHNALFFRSTIDDDVRIISFPISIV	71	
QY	57	-INVSFSDGIDGLOPGDDGKNRYSEKMSRKTRKLSOTLAANNVDKSKT-TEVYISH	114	
Db	72	TIKRPSAEVLEPLGIGISFNSSATS--PEMKRKFVYLS--LQKPMEEKIRRLRDCIFSD	127	
QY	115	L-FMAKAVAREHELSALTIMEPATVLDIFPFYNGYSDSIDAGSDAHLPGGLPVLAQ	173	
Db	128	MYFPWYTDIDELHIFRILYLNLSATWYCSIMH-----NAKYVRPHQPMUDE	174	
QY	174	RDPEFLLPSTHERF-----RSLMKKLETLLEGEEKPK--VLVNSFDA	214	
Db	175	SQ--SFVAVPLPEIRKFKLSOLTDLRKSDQKTYFDELLQVEDESEBSRGYIHDFTYE	232	
QY	215	LEP--DALKAIDKYEMIAIGPLISAPFLDCKDPDRSFGGDLPEKKSNDODCLEMSTN	271	
Db	233	LEPVIYVDYIYKRLKPKCHMGPRL--SHF-----ASKIRSKELLSEHNHNEIYIDWLNIAQ	284	
QY	272	PRSSVYVSFGSFVNTTKSQMEIARGLLDCGRPLAVVRVNE--GEEVLISCMEEIKRY	329	
Db	285	KPKSVLTVLSFGSGMARPEESOLNEIAQALDASNVNPIFLVLRNEETASMLPVGNLEDDTKK	344	
QY	330	G-K-LVKSQGLLEVLTNPSLGCFVTHCGMNSTLESISFVVPVAPAPWPFDDGNAKLME--	386	
Db	345	GLYIKGWNPDLTIMHSATGCFMTHCGNSTVLEALTIFGVPIATWPLVADQFVNEKRYEVR	404	
QY	387	-----DVMRTGVRVRANEGSVVVDGDEIRRCIEEYV--DGGEKSRKLRRESAGKKWL	436	

```

QY 1 MSRAHVLTLTFPGAGINPALOFAKRLANDIOYTFPTSVYARBRMSRAAOSNLIWF 60
Db 1 MKRAELVIITPFGIHSITGBELAKLAEHREHISIVILIKPNPTKV---SNLKL 56
QY 61 SFSDGDTODGLQREDDGKNYMSKSGIK-----ALSDTLAANNVDOK 10
Db 57 STA-----SRKIVIELKOETISITETGFLQKFEIESHKAQVDFLAGVACE- 103
QY 104 SSKITFVYVSHLPAMAKAREPFLRSALLIMEPAVLDFEYFENGYSO-----ETDA 15
Db 104 SVELSGVIAIDMCTSHIDYANFEVPSVFFTSAAMLGL-WHFGSLDFONGKIVYDKD 16
QY 158 GSDAHLPGGLPYLAORDLPSELLPSTH-----EFRSLMKKLETTLEGEKPKYL 208
Db 163 SETVLSIPAEONLVPAQVLPMEFIETEDGCCASADIGKRPR-----ETKGI 209
QY 209 VNSFDALPFDALKALDKYEM-----AIGPLIPSAFLDGKRPDSRSGSLFEKSGN- 266
Db 210 INFPLESHALEBSJSTDETPPYVTYGPPI-----LGRKSGSLESL 250
QY 261 -DDDCLEMLSTNPRSSVYVSRGSEFVNTKSKOMEIARGLDLCGRPFLVWVFN----- 313
Db 251 ETEKILKMLDMQPEKSVYFLCGSHGHEAQVKEIYVALESNGHFLMSLKKPPLOKF 310
QY 314 EG-----EYVL-ISCMBELKNGKVIYSMCSOLEVLTNPDSLOCFPYTHCGSNSTLESISF 365
Db 311 EGREXEYENLEBYLPEGFLEKRIANTGMVYGMAPQTAVLSHSAVGGFVSHCGMNSTLESITMF 370
QY 366 GVPVMAFEPOMEGOTNAKIMEDVYPTGQVRYAN-----EGSGVYDGEIRCEIE 415
Db 371 GVPVAMVLPFAEQONMA--FELVKEIAGLAEVYKMDYKKDKPNADDEIYRADIIEKIKI 422
QY 416 VMNGEKSRLKRESAGKMKDLARRKMEGDSVYNNLKVFLDEYV 459
Db 429 LMD---PENGIRKRYKEMKEVSRLAVEGSGSSSLKDFINDYI 469

```


Db 394 GDOAYNTNKYEEL 406

RESULT 15

R10429

ID R10429 standard; Protein: 506 AA.

DT 10-APR-1991 (first entry)

DE Ecdysteroid UDP-glucosyl transferase.

KM Ecdysteroid UDP-glucosyl transferase; EGT; insect control; polyhedrosis virus.

OS Autographa californica nuclear polyhedrosis virus strain IL.

PN MO9100014-A.

PD 10-JAN-1991.

PE 28-JUN-1989; U03758.

PA (UYGE-) UNIV GEORGIA RES FO.

PI Miller LK, Orellly D;

DR MPI; 91-036527/05.

DR N-PSDB; Q10336.

PT Biological insect control agents - comprising nuclear polyhedrosis virus with inactivated EGT gene encoding ecdysteroid UDP-glucosyl transferase

PS Disclosure: Page 14-14/1; 64pp; English.

CC The protein affects the growth, development or behaviour of insects.

CC The gene encoding the protein is either inactivated to prevent

CC insect moulting and pupation or is inactivated to reduce the

CC feeding behaviour, inhibit growth and result in the earlier death

CC of the insect host.

CC Pref. the egt gene is inactivated by replacing a portion of the gene

CC with a bacterial sequence encoding beta-galactosidase (VEGTZ); or

CC part of the gene is deleted without replacement (VEGTDEL).

CC A (recombinant) ecdysteroid UDP-glucosyl transferase having this

CC amino acid sequence, or a sequence about 70% homologous,

CC is claimed (52-53).

CC Sequence 506 AA;

SQ

Query Match 5.8%; Score 140.5; DB 1; Length 506;

Best Local Similarity 22.7%; Pred. No. 2.6e-06; Matches 98; Conservative 68; Mismatches 146; Indels 119; Gaps 24;

```
QY 29 NADIQVTFTSVYAMRMSR-----TAAGSNGLINEVSPSGCYDGLQPGDGK 77
Db 73 NADMSVQYKKLVANSAMFRKRGVSDTDTVTAAANYGLIEM--FKDQFDN----- 121
QY 78 NYNEMKSRGKIALSDTLAANNVDKSSKITVYVSHLEFAMAARAREFHLSALMLNIEP 137
Db 122 ----INVRLLINNOTFDLVVEAFADYA--LVFGHLY-----DP 155
QY 138 ATVLDPYFFENGYSDEID-AGSDAHLPGGLPVLAQRDLPSFLPSHERFRSLMKRKL 196
Db 156 APVYQIAPGY--GLAENPDYGAVARH-----PV-----HHPIMRSNEDDTEANVME-- 202
QY 197 ETLGEKEKPVLYVNSFDAL-----EPDAKAKIDKEMTAIG--PL-----IPSAFL 240
Db 203 --MKLYEKFELNMSNALKQOFGPNTPTIEKLKNKVOULLNLHPLEDDNNRPV--- 256
QY 241 DGKPSDRSFGDLFEKSGNDCCLEMLS-----TNPSSVYVVSFGSFVNTTKSQME 293
Db 257 ----PSVOYLGGGHLVKSAP--LTKLSPVINAQMNNKSKGTIVYVSGSID--TKSFAN 308
QY 294 EIAKGILLDCGR-----PFLMVYVYNGEEVLISCMEELKRVGKIY--SMCSQLEVLTHPS 346
Db 309 EELMYLINTFKTLDNNTILMKI---DEV---VKNTLPLPANNVTQNMFWNORAVLRHK 360
QY 347 LGCEVTHCGNNTLESISGCVPAVAFPOWFDQGTNAKLMEVDWRTGVVRANEGSVYDG 406
Db 361 MAAPITGGIGSSDEALMAGIAPVCLPMNGDQFYHAHKLQOL--GV--ARALDT--VTYSS 415
QY 407 DEIRCIIEEYV 417
Db 416 DQELVAINDVL 426
```

Search completed: August 1, 2000, 19:50:22
Job time: 21502 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:31 ; Search time 59.57 Seconds

(without alignments)
118,629 Million cell updates/sec

Title: US-09-147-955-6

Perfect score: 2407

Sequence: 1 MSRAHVLAATFPAGHINPA.....EDGSSVNLKVFLEDEVGCI 461

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: ./cgn2_6/pdata/2/1aa/5A.COMB.pep:*
2: ./cgn2_6/pdata/2/1aa/5B.COMB.pep:*
3: ./cgn2_6/pdata/2/1aa/6.COMB.pep:*
4: ./cgn2_6/pdata/2/1aa/PCYUS.COMB.pep:*
5: ./cgn2_6/pdata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	751.5	31.2	471	2	US-08-466-583-2
2	751.5	31.2	471	4	PCT-US95-07820-2
3	430	17.9	471	3	US-09-106-464-2
4	405.3	16.8	488	2	US-08-797-228-2
5	390	16.2	131	2	US-08-466-583-5
6	390	16.2	131	4	PCT-US95-07820-5
7	267	11.1	131	2	US-08-466-583-6
8	267	11.1	131	4	PCT-US95-07820-6
9	227	9.4	63	2	US-08-466-583-8
10	227	9.4	63	4	PCT-US95-07820-8
11	154.5	6.4	529	4	PCT-US92-00282-7
12	146	6.1	531	4	PCT-US92-00282-7
13	140.5	5.8	506	5	PCT-US92-00282-6
14	136.5	5.7	414	2	US-08-750-524-1
15	134	5.6	531	4	PCT-US92-00282-5
16	129	5.4	533	4	PCT-US92-00282-3
17	129	5.4	534	4	PCT-US92-00282-4
18	125	5.2	56	2	US-08-466-583-4
19	125	5.2	56	4	PCT-US95-07820-4
20	110	4.6	58	2	US-08-466-583-9
21	110	4.6	58	4	PCT-US95-07820-9
22	105	4.4	52	2	US-08-466-583-7
23	105	4.4	52	4	PCT-US95-07820-7
24	103	4.3	74	4	PCT-US92-00282-24
25	99.5	3.9	1258	2	US-08-310-912A-107
26	94	3.1	431	1	US-07-783-705A-2
27	91.5	3.8	481	1	US-08-489-733-3
28	91.5	3.8	491	2	US-08-993-581B-3

29	90.5	3.8	419	3	US-08-115-753-2	Sequence 2, Appl
30	90.5	3.8	419	3	US-08-115-753-33	Sequence 33, Appl
31	88.5	3.7	413	2	US-08-282-197C-49	Sequence 49, Appl
32	88	3.7	365	1	PCT-US93-03077-1	Sequence 1, Appl
33	85.5	3.6	1093	4	PCT-US93-03077-1	Sequence 1, Appl
34	84	3.5	816	1	US-08-190-802A-54	Sequence 54, Appl
35	83	3.4	804	2	US-08-785-428-2	Sequence 2, Appl
36	83	3.4	804	2	US-08-996-797-2	Sequence 2, Appl
37	82.5	3.4	409	2	US-08-924-254-2	Sequence 2, Appl
38	82.5	3.4	409	2	US-09-120-249-2	Sequence 2, Appl
39	82	3.4	330	3	US-08-115-753-1	Sequence 1, Appl
40	81.5	3.4	855	2	US-08-468-558-2	Sequence 2, Appl
41	81	3.4	821	1	US-07-928-464-2	Sequence 2, Appl
42	81	3.4	821	1	US-08-003-311B-2	Sequence 2, Appl
43	81	3.4	821	1	US-08-261-432-2	Sequence 2, Appl
44	81	3.4	821	4	PCT-US93-07347-2	Sequence 2, Appl
45	80	3.3	399	1	US-08-096-623A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-466-583-2
; Sequence 2, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szczyglowski, Krzysztof
; APPLICANT: Szczyglowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-583-2

Query Match 31.2% Score 751.5; DB 2: Length 471;
Best Local Similarity 38.18; Pred. No. 11e-69;
Matches 182; Conservative 75; Mismatches 180; Indels 41; Gaps 13;

```

OY 5 HVLATFPAGHINPALOFAKRLANADIQVTFSTVYAMRMRSTAAGSNGLINVFESD 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 HVLVFPFGCGHNMVPOFAKRLASKGVATTLVTT---RFIORTADVDAHPAWVAISD 59
OY 65 GYDDGLOPDDCKNNMSMKSRGKIALSDTLAANNVDOKSSK---ITFVYSHLFMAAK 121
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 GHDEG---GFASAAGVAYLEKQAAASASL-ASLVEARRASADATFCVYDSYEDWVLP 115
OY 122 VAREFLRSALLMIEPATVLDIFFFYFNG-----YSDETAGSDAHLPG---GLPV 170
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 116 VARRMGLPAVPSTOSCAVSAVYHFSGRLAVPAGAAADSGGAGAAALSEAFGLPE 175
OY 171 LAORDLPFLPSTHERPRLMKKEKLETFEGEEKPK-VLVNSPDALPOLKAIDKY-EM 228
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 MERSELPSPVFP--DHGPTPTAMQAIKOFAGAKDWDVLFNSFELETFEVLGLTYILKA 233
OY 229 IATGLIP-----SAFLDGKDPDSRFGDLFEKGSNDDCLEMLSTNPRSSVYVSFG 282
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 RAIGPCVPLPTAGRTAGANGR---ITYGANLVKP---EDACTKMLDTKPDPSVAIVSFG 286
OY 283 SFVNTKSGMEIARGLDCGRPLMVYRVNVEGEEVLISCMELKRYG--KIYWSCQLE 340
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 287 SLASLGNACKKEIARGLLAAGKPLMVYRASDEHQVRYLLAERTATGAANVYVPCPOLD 346
OY 341 VLTBPSLGFVTHCGMNSTLESISFGVPVNAFPQWPDGCTNAKLMEDVWRTGVRVRAEE 400
Db 347 VLAHPVAGCFVTHCGMNSTLESISFGVPVNAFMLTQPTNANVELANGAGARARADAG 406
OY 401 GSIVDDDELRCIEEYMDGGEKSRRLRESAGKMDLARKMEBGSVNLYKFLEDEY 458
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 407 AGVFLGVEVERCVRAVMDGGEAASAKRAAGEMWDRARAVALPAGCGSSDRNLEDFEYGV 464

```

RESULT 2

```

PCT-0595-07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szezygowski, Jerzy B.
; APPLICANT: Szezygowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Wanner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-0595-07820-2

```

```

Query Match          31.2%; Score 751.5; DB 4; Length 471;
Best Local Similarity 38.1%; Prod No. 1.1e-69;
Matches 182; Conservative 75; Mismatches 180; Indels 41; Gaps 13;

```

```

OY 5 HVLATFPAGHINPALOFAKRLANADIQVTFSTVYAMRMRSTAAGSNGLINVFESD 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 HVLVFPFGCGHNMVPOFAKRLASKGVATTLVTT---RFIORTADVDAHPAWVAISD 59
OY 65 GYDDGLOPDDCKNNMSMKSRGKIALSDTLAANNVDOKSSK---ITFVYSHLFMAAK 121
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 GHDEG---GFASAAGVAYLEKQAAASASL-ASLVEARRASADATFCVYDSYEDWVLP 115
OY 122 VAREFLRSALLMIEPATVLDIFFFYFNG-----YSDETAGSDAHLPG---GLPV 170
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 116 VARRMGLPAVPSTOSCAVSAVYHFSGRLAVPAGAAADSGGAGAAALSEAFGLPE 175
OY 171 LAORDLPFLPSTHERPRLMKKEKLETFEGEEKPK-VLVNSPDALPOLKAIDKY-EM 228
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 MERSELPSPVFP--DHGPTPTAMQAIKOFAGAKDWDVLFNSFELETFEVLGLTYILKA 233
OY 229 IATGLIP-----SAFLDGKDPDSRFGDLFEKGSNDDCLEMLSTNPRSSVYVSFG 282
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 RAIGPCVPLPTAGRTAGANGR---ITYGANLVKP---EDACTKMLDTKPDPSVAIVSFG 286
OY 283 SFVNTKSGMEIARGLDCGRPLMVYRVNVEGEEVLISCMELKRYG--KIYWSCQLE 340
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 287 SLASLGNACKKEIARGLLAAGKPLMVYRASDEHQVRYLLAERTATGAANVYVPCPOLD 346
OY 341 VLTBPSLGFVTHCGMNSTLESISFGVPVNAFPQWPDGCTNAKLMEDVWRTGVRVRAEE 400
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 347 VLAHPVAGCFVTHCGMNSTLESISFGVPVNAFMLTQPTNANVELANGAGARARADAG 406
OY 401 GSIVDDDELRCIEEYMDGGEKSRRLRESAGKMDLARKMEBGSVNLYKFLEDEY 458
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 407 AGVFLGVEVERCVRAVMDGGEAASAKRAAGEMWDRARAVALPAGCGSSDRNLEDFEYGV 464

```

RESULT 3

```

US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 6011145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Changas, Guider S.
; APPLICANT: Kuhl, Jian Ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc-6-Patty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,464
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,554

```

FILING DATE: 13-AUG-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Spectol, Eric S.
 REGISTRATION NUMBER: 22495
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-415-1508
 TELEFAX: 703-415-1508
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 471 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-106-464-2

Query Match 17.9%; Score 430; DB 3; Length 471;
 Best Local Similarity 28.7%; Pred. No. 3.3e-36;
 Matches 146; Conservative 72; Mismatches 197; Indels 94; Gaps 19;

QY 1 MSRAHVLATFPAGHINPALQFAKRLANADIQVTFEY---SVYAMRRSRRTAASNG 55
 DB 2 MGLHPEFFPMAGHNPILDMAKIVACRGVAKITITPLNESVFSKAIERNHIGIEI 61
 QY 56 LINFVSSDGIID-----GLQPGDDG-KNYMSEMSRGIKALSDTLAANNVDQSSK 106
 DB 62 DIRLKPFAKENLPEDCERLDVPSDDKLPFLK-----AAAMMDDEEEL 108
 QY 107 I-----TFVYVSHLFAMAARVAREFRLRSALMIEPATYLDIEFYFNGYSDSIDAGSDA 161
 DB 109 IGECRPPCLVSDHFLPWTIDSAKFSI-----PRIV-----FICSTYFALCYGDS 153
 QY 162 IHLPGILVLAQDRLPSLIPSTHEREFSLMKEKLETLGEEKPK----- 206
 DB 154 IRRKKPKNVSS-DTEFFVVPDPFHEIR-LTRTOLSPFQSDPEQGMAPKIKAVRESPDA 211
 QY 207 ---VLVNSPALPPDAIKADIKYEMI-----AIGPLIPSAFLDKDSDSPGDDLPFEK 257
 DB 212 SYGVIRFNSFIEESD---YEHTTKYVGRKNNAIGPL-----SLCRNDIEYAAENG---RK 261
 QY 258 GSNDND-CLEMLSTNPRSSVYVSGSFYNTKSSQMEIARGLDCCGPPLMVVYVNEGE 316
 DB 262 SSIDERRACLMLDSKSSSIYVYVCGSTADFTTAOMOBELAMGLESGDDFTIWIIRTGNED 321
 QY 317 EVLISCEELKRVGKIY-SWCSOLEVLTHPSLGCFTYHCGNSTLESISPCVPVAPAPW 375
 DB 322 WLPGEFERKREKGLIIRGMARVYLIDHEIAGATYTHCGNSTLEISAGVPHLTPVYF 381
 QY 376 FDQGTNAKLMEVYRTGVYV-----RANEESGVVDGDIRRCIEEYVDGSEKSRKLRES 429
 DB 382 AEQFENKELTYEVRSGAGVSGKQMRRTASEG--VKREAIATAIKRVA-ASEETEGFRSR 438
 QY 430 AGKWKDLARAMEEDGSSVNNLKVFLDEV 458
 DB 439 AKKKEMARARAIPEGSSYNGWATLIQDI 467

RESULT 4
 US-08-797-226-2
 Sequence 2, Application US/08797226
 Patent No. 5959180
 GENERAL INFORMATION:
 APPLICANT: MOERS, CHARLES P
 APPLICANT: ALLEN, PAUL V
 APPLICANT: ROCKHOLD, DAVID R
 APPLICANT: STAPLETON, ANDREW
 APPLICANT: GABBARINO, JOAN E
 APPLICANT: FRIEDMAN, MENDEL
 APPLICANT: BELKNAP, WILLIAM R
 TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
 TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
 TITLE OF INVENTION: GLYCOLALDEHYDES IN SOLANACEOUS PLANTS
 NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
 ADDRESSEE: NANCY J. PARSONS
 STREET: 800 BUCHANAN ST.
 CITY: ALBANY
 STATE: CA
 COUNTRY: USA
 ZIP: 94710
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/797,226
 FILING DATE:
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: PARSONS, NANCY J
 REGISTRATION NUMBER: 40,364
 REFERENCE/DOCKET NUMBER: 0011.97
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 559-5773
 TELEFAX: (510) 559-5777
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 488 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-797-226-2

Query Match 16.8%; Score 405.5; DB 2; Length 488;
 Best Local Similarity 26.1%; Pred. No. 1.2e-33;
 Matches 131; Conservative 93; Mismatches 199; Indels 79; Gaps 18;

QY 5 HYLATFPAGHINPALQFAKRLANADIQVTFEY---AVRRMSRRTAASNGI----- 56
 DB 12 HYLPLPISAGHPFLPVNMAARLPRASRGVAKITITTHNALLRSTIDDDVRLSGPISIV 71
 QY 57 -INVEYSDGYDDGLQPGDDGKNYMSKSRGIKALSDTLAANNVDQSSKI-TFVYSH 114
 DB 72 TTKFSAEAVGPEGIESFNSGTS--PEMPKIFVALS--LQKPMEDKRLRELRPCIFSD 127
 QY 115 L-FAMAARVAREFRLRSALMIEPATYLDIEFYFNGYSDSIDAGSDAHLPGGLPVLAQ 173
 DB 128 MFPWTVIADIELHPIRLNLSAKYCSIMH-----MLKYRPHKQPNIDE 174
 QY 174 RDLPSFLIPSTHEREF-----RSLMKEKLETLGEEKPK--VLVNSFDA 214
 DB 175 SQ--SFVVPGLPDEIKFKLSQTLNDDLRKSDQKTVDELLQGVESSEERSYGIHDTPEY 232
 QY 215 LEP--DALKAIDYEMAIAGPLIPSAFLDGNPDRSGFLFEKGSNDCCLEMLSTN 271
 DB 233 LEPATVDTYQKLKPKCHRGPL--SHF-----ASKISKELISEHNNELVIDMLNAQ 284
 QY 272 PRSSVYVSGSFYNTKSSQMEIARGLDCCGPPLMVVYVNE--GEEVLLSCMEELRV 329
 DB 285 KRKSALYVSGSMARFESQLEINELALDASNVPLFTFVLRPNDEETASWLPVGNLDDKTKK 344
 QY 330 G-KIYVSMCSOLEVLTHPSLGCFTYHCGNSTLESISFGVPVAPAPWOMFOGNNATIME-- 386
 DB 345 GLIYIGWVPOLITMEHSGTFMHCCTNSVLEALITFGVPMITPXYADQFNEVVEYR 404
 QY 387 -----DWRGTGVRRANEEGVVDGDIRRCIEEVM--DQGEKSRKLRESAKRWDL 436
 DB 405 GLIGIGIDVNNEGIEI---TGPIVESAKIRIARILMINSGEETIINIRDVVAMSKM 460
 QY 437 ARKAMEEDGSSVNNLKVFLDEV 458
 DB 461 AQNATINEGSSMNNLTALQIH 482

Page 5

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07820
FILING DATE: 19-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,427
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 11-94B PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TELEX: 49617834
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 131 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
PCT-US95-07820-6

Query Match 11.1%; Score 267; DB 4; Length 131;
Best Local Similarity 41.2%; Pred. No. 3.5e-20;
Matches 54; Conservative 16; Mismatches 59; Indels 2; Gaps 1;

QY 264 CLEWLTNPRSSVYVYSGSPFVNTKSGMEIARGLLDGRPLVAVRVNDEGEVLSICM 323
Db 1 CLMLRQRPARGAAYVSEFVACPPDRLRELALGLEDGAPFLMSIKRDSWPHLPGL 60
QY 324 ELKRVK--IYSMCSOLEVLTHPSLGCFTVHCGMNSTLESISFGVPMVAFPQMDQGTN 381
Db 61 DRAGAGSGSLVYPMAPQVAVLHRPSVGAFTVHAGMASVLEGLSSGVPMACRPFGDORNN 120
QY 382 AKLMEDVWRTG 392
Db 121 ARSYAHVWGFG 131

RESULT 9
US-08-466-583-8
: Sequence 8, Application US/08466583
: Patent No. 5919998
GENERAL INFORMATION:
APPLICANT: Bandurski, Robert S.
APPLICANT: Szeleszen, Jedyzej B.
APPLICANT: Szczygowski, Krzysztof
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
TITLE OF INVENTION: and Plant Growth.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee & Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,583
FILING DATE: 06-JUN-1995

```

CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,427
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Felder, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 11-94A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: Internal
US-08-466-583-8

Query Match 9.4%; Score 227; DB 2; Length 63;
Best Local Similarity 69.1%; Pred. No. 1.5e-16;
Matches 38; Conservative 8; Mismatches 9; Indels 0; Gaps 0;
Qy 332 IVSWCSQLEVLTHPSLGCFTVHCNMSTLESISFGVPMVAFPQWFDQGTNAIME 386
Db 9 VVPCRPQDLVLAHPVAGCFVTHCGMNSTLEALSFGVPMVAMALMTDQPTNANVE 63

RESULT 10
PCT-US95-07820-8
Sequence 8, Application PC/TUS9507820
GENERAL INFORMATION:
APPLICANT: Bandurekl, Robert S.
APPLICANT: Szerzen, Jędrzej B.
APPLICANT: Szczygielowski, Krzysztof
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
TITLE OF INVENTION: and Plant Growth.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee & Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07820
FILING DATE: 19-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,427
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Felder, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 11-94B PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: Internal
PCT-US95-07820-8

Query Match 9.4%; Score 227; DB 4; Length 63;
Best Local Similarity 69.1%; Pred. No. 1.5e-16;
Matches 38; Conservative 8; Mismatches 9; Indels 0; Gaps 0;
Qy 332 IVSWCSQLEVLTHPSLGCFTVHCNMSTLESISFGVPMVAFPQWFDQGTNAIME 386
Db 9 VVPCRPQDLVLAHPVAGCFVTHCGMNSTLEALSFGVPMVAMALMTDQPTNANVE 63

RESULT 11
PCT-US92-00282-7
Sequence 7, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OMENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-7

Query Match 6.4%; Score 154.5; DB 4; Length 529;
Best Local Similarity 22.5%; Pred. No. 1.9e-07;
Matches 68; Conservative 50; Mismatches 139; Indels 45; Gaps 10;
Qy 137 PAVTVIDFVFYNGYSDSIDAGDAIHLPGSLPYLAQRDPSFLPSTHERFRSLMKEXL 196
Db 186 PVSIVPRFTRK-----SDHMTFQRIANFINIENILYHCLSKYEILASDL 235
Qy 197 ETLEGEKRYLVNSFDALPDALKAIDKYEML-AIGPLIPSAFLDGGDPDSRSGGDL 254
Db 236 -----KRDV--SLPALHONSIMLL-RYDFVFEYPRPVPMPKMFIFG-----GTWC 276

Fri Aug 4 15:19:50 2000

us-09-147-955-6.rai

Page 9

Oy	421	EKSRK	425
	:		
Db	431	DKSYK	435

Search completed: August 1, 2000, 19:51:33
Job time: 18755 sec

100% 100% 100% 100% 100%

THIS PAGE BLANK (USPTO)

FEATURES	SOURCE
0y 1014	ggaaatcgtatcttggtttcttcaatggaaggtctgcagcactcccttggttgatgtt 1073
Db 471	GGAAATAGTACCATGATGATGTGTACAACTGTGAAGTCTTCACATCCATCTTTTGAAGATTT 530
0y 1074	tcgtacacacactgcgggttgaaattcgcactctagagaatatactcttcgggggtccgaatg 1133
Db 531	TTTCCTCCTACCTGTGATGGATGGAAATTCGACATCTTGAAAGCCATCTGCTCAGGCTACGTGG 590
0y 1134	tgtgtcttcgcgagtggttcgtatcaaggagcgaatgcgaagctatgagatgctt 1190
Db 591	TGGCATTTCTCTCATTTGACAGATTCAGGAGCATTAATGCMAAAGTATGTAAGAGATGTTT 647
RESULT 2	
AM349414/c	
LOCUS	747 bp mRNA EST 01-FEB-2000
DEFINITION	GM210007A20D2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
ACCESSION	AM349414
VERSION	AM349414.1 GI:6847124
KEYWORDS	EST.
SOURCE	soybean.
ORGANISM	Glycine max
REFERENCE	Elkayorta, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
AUTHORS	1 (bases 1 to 747)
TITLE	Vodkin, J., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V., Expelling, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H. A Functional genomics Program for Soybean (NSF 9872565)
JOURNAL	Unpublished (1999)
COMMENT	On Jul 9, 1999 this sequence version replaced g1:5434916. Contact: Vodkin, J.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
	Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
	University of Illinois
	Edwin R. Medisan Building, 1201 W. Gregory, Urbana, IL 61801, USA
	Tel: (217) 244-6147
	Fax: (217) 333-4582
	Email: l-vodkin@uiuc.edu
	This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888)919-3524 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
	Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
	Location/Qualifiers
	1..747
	/organism="Glycine max"
	/cultivar="Williams"
	/db_xref="taxon:3847"
	/clone_lib="Gm-r1021"
	/tissue_type="root"
	/lab_host="X110-Gold"
	/note="vector: pbluescript II XR. Site.1: EcoRI, Site.2: XhoI. Library Gm-r1021 is a sequence-diven, retracted set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. StrataGene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@nau.edu, virginia.h.coryell@nau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota.
	http://www.cbc.umn.edu/Research/projects/soybean/index.html
	. Retracting was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics,

BASE COUNT	206 a	235 c	97 g	187 t	22 others
ORIGIN					
Query Match	12.0%	Score 248.2	DB 71	Length 747	
Best Local Similarity	62.7%	Pred. No. 2,46-45			
Matches 427	Conservative	0	Mismatches 233	Indels 21	Gaps 3
Oy	728	atccctccgcatctcttggagcggtaaaagatccctcgcagcaggctcttcgcgcgagatttg	787		
Db	733	ATTCTNNNNNNNCNTTNNNNNNCGTAAAGATCCCGGGATCTACTNNNTATGCGGTGTATNN	674		
Oy	788	ttcgagaagaagtcgaatgacgacatgacctgcgaatcgatgttgagcaagaatccgcgatc	847		
Db	673	NNC-----NNCTCTCAATGATTATTCGTTGAATGCTTGAGATCGTTGCATCAGCCTGAGTTA	623		
Oy	848	tcggtggttaacgttcctcgttcggagagcttcgtttaacgaagaagctcgcaatcggagaag	907		
Db	622	TCGTGGTTATATTTTCATTTTGTAACCTTCCTGCTGTGCTGTGATACAGATGAAGAGG	563		
Oy	908	atagaagaagagcgtctgaattctgtggagagccgtttctgttggcgtaagaagtaaacgaa	967		
Db	562	CTTGACAGCGCGCTCGTGAATCCGATATCTCTTCTTGCGGTCAATTAGATGATGCA	503		
Oy	968	gggagaagaagttatgataagttgcattcgatlgagagagtttgaagaagctggggaaatgtatct	1027		
Db	502	GGATATGAAG-----ATTAATGCAAGAGAGGAACTGGAGCAGAGGGGTAAATGTGAAA	449		
Oy	1028	tgtgtctcaatttgaagctcctgaacgcacccctcgtttggagatgtttcgtgcacacac	1087		
Db	448	TGCTGTTCTCAGGTGGAGGTCTGTGCGCATGGTTCCCTGGGTTGTTTGTAAACGCAATGT	389		
Oy	1088	gggtggagattcgcctctagaagttatcccttcgggggttcgaatgggtgttttcgcgag	1147		
Db	388	GGATGCAATTCGACTATGAAAGATTGGGTTGGGGGTTCTTATGAGTGCGGTTTCCCGAG	329		
Oy	1148	tggtctcacaagaagagcgaatcgaagaactgatatgagagatgtgtgagagccgggtgtgaa	1207		
Db	328	TGGACACACACCAAGCGACGAATGCAAAAGATGGTCACAGATGTGTGAAAGACGGGGGTGAG	269		
Oy	1208	gtg-----agagctaatgagagaggttagcgctgtgtgtgtgtgtgaatttagagaigt	1261		
Db	268	GTCGATGATATAGTGAATGTGTGACGAAGGATATGTGAAACGACAGAGGAGATATAGCAATGT	209		
Oy	1262	atggaagaggttatggaatggggagaaagaatgagaaacttagagaagctgtcggcaag	1321		
Db	208	TTGGATGTGCTCATGGGAGGTGAGAAAGAACACGGAATTCACAAAGAAATGCTGATPAAA	149		
Oy	1322	tggaagaagatttgcgaagaagaacttggagagagaatgatatagttacgttaaaccccaag	1381		
Db	148	TGGAATAGCTGTGGCCAGGGAAGCCGTGACAGGAAGAGAGGCTTTCGATATATATAGAG	89		
Oy	1382	gtcttcttgatgaggttgta	1402		
Db	88	ACTTTCCTCATGATGTGCA	68		
RESULT	3				
LOCUS	A1771830	536 bp	mRNA	EST	29-JUN-1999
DEFINITION	EST252950 tomato ovary, TAUU Lycopersicon esculentum cDNA clone				
ACCESSION	A1771830				
VERSION	A1771830.1	GI:5269871			
KEYWORDS	EST.				
SOURCE	tomato.				
ORGANISM	Lycopersicon esculentum				
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
	Magnoliophyta; eudicotyledons; Asteridae; eunasterids I; Solanales;				
	Solanaceae; Solanum; Lycopersicon.				
REFERENCE	1 (bases 1 to 536)				

AUTHORS Alcala,J., Vredalov,J., White,R., Matera,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Romling,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nieman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Mar 8, 1999 this sequence version replaced gi:4388463.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU

FEATURES

source
location/Qualifiers
1..536
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLE038G17"
/issue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 187 a 66 c 138 g 145 t
ORIGIN

Query Match 10.9%; Score 224.2; DB 44; Length 536;
Best Local Similarity 67.3%; Pred. No. 5,1e-40;
Matches 367; Conservative 0; Mismatches 163; Indels 15; Gaps 3;

748 cgttaaagatccttcgacacagcttcctcgcgagagatcttcgagaagaagctgaatga 807
1 CAGAAAATTCATTCGAAATCTTCATTTGCTGCTGATCTTTTCAAAA-----GTCAA 54
808 cgaagatcgcctgaatggttgagacgaatcctcgatccttggtgttaccgttcgtt 867
55 TGATGACTGATGGAATGCTTAAACACAAAGCCCTTAATCATCATTTGTTATATCTCAT 114
868 cgaagatcgcctgaatggttgagacgaatcctcgatccttggtgttaccgttcgtt 927
115 TCGGAGCTGATGGAATTTATCAAGAAACCAAGGAGGATTCGAAAAGGTTGATAGA 174
928 ttgtggagggcggcttttgggtgtgtgaag-----agtaaacgaaggaaggtctt 981
175 GATCCAAAGGCCATTTTATGCGTAATMAAGATCAAGAAAGAAAGAAAGAGAA 234
982 gataagtcgacgtgagagatgtaaacgagtggtggaataatgtatcttggttcacatt 1041
235 ATTAGTTCGATGATGGAATTTAGAGAACGAAGGAAATAGTATCCATGCTGTTCAACT 294
1042 ggaagatccttcgacatccctcggttggtgttcttcgagacacatcggttgaaatcga 1101
295 TGAATCTCTGACATCCATCTTTAAGATGTTTGTCTGCACTGCGATGGAATTTGAC 354
1102 tctaaagatcatalcttcgggtgttcgagatggtgttccttcgacgttgatcgaatcga 1161
355 TCTAGAAAGCCATTCGATGAGCTGTACTGTGTGGCATTTTCCATGACACAGATCAAG 414
1162 gacgaagtcgaagatcgtgagagatggtgttgagagacggtgtgagagagagcacaatga 1221
415 GACTATATCAAGATGATGGAATGTTGTAAGACAGAGGTGAGAGTACAGAGTAAAGA 474
1222 ggaagatcgcctgaatggttgatgaatacgaagagatgatgaagaggttatgatgatg 1281
475 AGATGGT---GTGGTTGAGAGTATGAATAAAGAGTGCATGAATAATTGATGATGG 531

QY 1282 gggag 1286
DB 532 TCGAG 536

RESULT 4
AI897621
LOCUS AI897621 546 bp mRNA EST 27-JUL-1999
DEFINITION EST267064 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
AI897621
ACCESSION CLED28K21, mRNA sequence.
VERSION AI897621
KEYWORDS AI897621.1 GI:5603523
SOURCE EST.
ORGANISM tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asterales; eumastids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 546)
Alcala,J., Vredalov,J., White,R., Matera,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Romling,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nieman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Apr 7, 1998 this sequence version replaced gi:3035329.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU

FEATURES
source
location/Qualifiers
1..546
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLED28K21"
/issue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 196 a 82 c 114 g 154 t
ORIGIN

Query Match 10.1%; Score 207.8; DB 45; Length 546;
Best Local Similarity 66.4%; Pred. No. 2,3e-36;
Matches 334; Conservative 0; Mismatches 157; Indels 12; Gaps 2;

600 tgaagagaatctggaatcttgaagagtgaaagaaacctaagcttctgtggaacagct 659
50 TCAAGACCACTAGACCATTTAGATGATGTAAGAAATCCAAAGGTACTTGAATTCAT 109
660 tctatgctctgagagctatgctcaagagccatgatgaatgagatgcatgcatcg 719
110 TTGATGATTTAGAGCTAGAGCCTCAACCTATTGAAAAATATCAATTTAATGGAATTG 169
720 ggcgctgattccttcgacatccttggaagtgtaagatccttcgacagagcttcggcg 779
170 GACATTTATTTCTTCAATCTTTGGTGGAAATTCATTTGGAATCTTCATTTGGTG 229
780 ggaatcttcgagaagaaggtcgaatgacgaacgacctcgatggttgacacgaatc 839
230 GTGATCTTTTCAAAA-----GTCAATATGATGATGATGATGATGATGATGATGATG 283


```

/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-blue MKR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni Laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      201 a      84 c      112 g      162 t
ORIGIN

```

```

Query Match      9.8%; Score 202.6; DB 45; Length 559;
Best Local Similarity 66.3%; Pred. No. 3.3e-35;
Matches 327; Conservative 0; Mismatches 154; Indels 12; Gaps 2;

QY 600 tgaagagaatttggaaactttaagaagtgagaagaaactaaagcttgtaacagct 659
DB 73 tcaaaagaccactagaccattagatgataagaaagaaagctgacttggaatgacat 132
QY 660 ttgatgcgttgagcctatgagcctcaagaccattgataagtaagagatgattgacatcg 719
DB 133 ttgatgacattagagctagaccactcaaacctattgaaaaatgcaatttaattggaattg 192
QY 720 ggcgcgtgattccctccgcatcttggaagcgttaagatcccttcggacaggtcttcggcg 779
DB 193 gaccattgattcttccatcattcttggtgtaaaagattggaattcttcatatttggtg 252
QY 780 ggaattcttcgagaaagggcgaatgacagcagattgctccgaatggttgagcaagaaac 839
DB 253 gtagcttttttcaaaa-----gtcaaatgatgactgacatgaaagtttaaacacaaagc 306
QY 840 ctccgctcttcggttggtttagcttgcgttcggaagcttcgttaataagaaagtcgcaaa 899
DB 307 cttaattcattcaattgttttatctcattttggagacttattgaaatttatcaaaacaa 366
QY 900 tggaaagagatgcgaagagggcgtctgtagatctgagaaagccgttcttggtggtgaag-- 957
DB 367 agagagagatgacaaagaggggtgatatgagatccaaagggcattcttatggttaataagag 426
QY 958 -----agtaagaaagagagaaagagatgataagttgacagggaggggttgaaagagagg 1013
DB 427 atcaagaagaagaaagaaagaaagaaagaaatttaagttgcatgattgaaagaaagcaag 486
QY 1014 ggaagaaatgtatctctggtctgctcaattggaagtcctgacgcaaccctcgttggaatgt 1073
DB 487 ggaagaaatgtacatgctggtgtcacaacttgaaagtcctgacacatccatttttgaagatgat 546
QY 1074 tctgagacagactg 1086
DB 547 ttgtctccgactg 559

RESULT 7
LOCUS AM442098 649 bp mRNA EST 14-FEB-2000
DEFINITION ESTJ11494 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
ACCESSION AM442098
VERSION AM442098.1 GI:6977349
KEYWORDS EST.
SOURCE
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 649)
Alcala,J., Vredalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upson,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,
Romling,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato fruit tissue

```

JOURNAL Unpublished (1999)
ON JUL 7, 1999 this sequence version replaced g1:5407380.

COMMENT

Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU

FEATURES

source
Location/Qualifiers

```

1..649
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEJ21G12"
/clone_lib="tomato fruit red ripe, TAMU"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopen accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe), 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT      217 a      118 c      124 g      190 t
ORIGIN

```

Query Match 9.3%; Score 191.6; DB 72; Length 649;

Best Local Similarity 59.9%; Pred. No. 9.5e-33;

Matches 392; Conservative 0; Mismatches 229; Indels 33; Gaps 3;

```

QY 345 tcaagttcgttggtactaccacacctcttgatbgygcgaagtgagcgtgagttcc 404
DB 2 tcaagtccttcgctgacacactctctgacacttgagcagctgagcgcgaactcc 61
QY 405 atctccgagagcgcctactctcgattgagtaagcagtaagtggttgatatattcttct 464
DB 62 atatccatcccgcttactatgattgattcaaccagcagactgctgacatattactattt 121
QY 465 attcaagcgtctatagcagcaaat-----cgatgcygggttcgagatctatc 512
DB 122 acttcaatgcatgagatgaaatgaagtgatgactcaaatgattccaaattggagta 181
QY 513 acttgccggagagactccagtgctgcccgaagtgatgttaacgcttccctccctcct 572
DB 182 tccaatggccaaagccttccattactaaaaagccaaagatcttccattttttttagtttct 241
QY 573 c-----cagcaatgaagatccgttcaactgaltgaagsgaaattgga 617
DB 242 ctagctcaaaagatgataagattgatttggctctaccacatttcaaaagcgaactagaca 301
QY 618 cttagaagtgtaagaaacctaagctcgttggtgaaacagcttgatctgttgagccg 677
DB 302 cattagatggttaaaaaaaatccaaagctggtgaaatgatttgatgcatgaaactgag 361
QY 678 atgagctcaagagcattgataagtaagagatgaltgcaatcgagccgttgatccctcg 737
DB 362 agcactcaaaagctattgaaaaaatacaatttaattggaattggaacattgattccctcat 421
QY 738 catctcttgagtgtaaaagatccctcgagacggtcttcctcgagagattgctcggaag 797
DB 422 cattcttggtggaaaaatattcattggaattcttatttttggtgctgattcttttcaaaa-- 479
QY 798 ggtcgaatgacagcagatgctcgaatggttgagcaagaaatccctcagatcgttggttt 857
DB 480 -----gtcaaatgatgactgacatgaaatggttaaacaaagccttaaatcatcatgttt 535
QY 858 acgttctcgttggaagctcgttaatacagcaagtcgcgaatggaagagatagacaag 917

```



```

Db 536 ATACCTATTGGAGCTTATGAATTTATCAAGAAACCAAGAGAGATTCCAAAG 595
QY 918 ggcctgacatgtg99agccgttttctg99gtgtaagcaacgaag 971
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 GGTGATGAGAGATCAAAAGCCATTCTTATGGGTAAATAGAGATCAAGAAAG 649

RESULT 8
A1487571 520 bp mRNA EST 29-JUN-1999
LOCUS EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLED10D21, mRNA sequence.
ACCESSION A1487571
VERSION A1487571.1 GI:4382942
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 520)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Niernan,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
JOURNAL On Apr 7, 1998 this sequence version replaced gi:3035667.
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.
LOCATION/Qualifiers
source 1..520
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED10D21"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED1 tomato Carpel EST library. Oligot-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 194 a 76 c 107 g 143 t
ORIGIN

Query Match 9.2%; Score 189.8; DB 40; Length 520;
Best Local Similarity 66.1%; Pred. No. 2,3e-32;
Matches 310; Conservative 0; Mismatches 147; Indels 12; Gaps 2;

QY 600 tgaagagaagaattggaacttagaaggtgaagaaacctaaggtctgtgtaagcgt 659
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 TCAAGAGCAACTGACACATTGATGCTGAAGAAATCCAAAGCTACTTGTGAATACAT 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 660 ttgatgcgttggagcctgtagcgctcaagggcaatgataagtagagatgcacatcg 719
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 TTGATGATGATGAGACTGAGACCTCAAGCTATTGMAAAATCAATTTAATGGAATG 176
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 720 ggcgctgattccttcgcacatcttgagcgtaagatccttcgacagcgtcttcgcgcg 779
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 GACCATGATTCTCTCATCTTCTTGGGTGGGAAAGATTCATGGAATCTTCATTTGGTG 236
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 780 gagatctgtcggaaagggctgaatgacagcattgtcctgaatgtgttgagcaagaac 839
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 237 GTGATCTTTTCAAA-----GTCAAATGACACTACATGGAATGGTTAAACACAAAGC 290
QY 840 ctgcattcgcgttgtagttagcttgcgttcggaagcctcgttaatacagcgaagtcgcaaa 899
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 CTAAATCATCAATGTTTATATCTCATTTGGAGTGCTATATGAATTTATCAAGAAACCAA 350
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 900 tggagagatagcagaagggcgtctgattgtgtggaagccgcttttctgtgtgtaag- 957
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 AGGAGGAGATTGCAAAAGGGTTGATGAGATCCAAAGGCCATTCTTAATGGAATTAAGAC 410
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 958 ----agtaacgaaggaagaggtattgtaagttcagtgaggagttgaaacgagtcg 1013
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 ATCAAGAAAGAAAGAAAGAAAGAAAGAAATTAAGTTGCATGATTCGATTAAGAAAGCAAG 470
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1014 ggaagaatgtatcttgcgttgcataatgtgaagctcgcagcattccctc 1062
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 GGAATAATGATCATGCTGTTCACAACTGGAATCCCTGACATCATC 519
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AM034683 685 bp mRNA EST 15-SEP-1999
LOCUS EST278414 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLED11D5 similar to UDP-glucose:anthocyanin
5-O-glucosyltransferase, putative, mRNA sequence.
ACCESSION AM034683
VERSION AM034683.1 GI:5893439
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 685)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
JOURNAL On Apr 3, 1998 this sequence version replaced gi:3018868.
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
LOCATION/Qualifiers
source 1..685
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED11D5"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLED - Cotyledons
or seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST library"

BASE COUNT 219 a 128 c 137 g 201 t
ORIGIN

Query Match 8.9%; Score 183.4; DB 63; Length 685;
Best Local Similarity 59.9%; Pred. No. 6.4e-31;
Matches 379; Conservative 0; Mismatches 221; Indels 33; Gaps 3;

```


Tanksley, S. D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:6677182.
Contact: David Fritsch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@CLEMSON.EDU

FEATURES
source

Location/Qualifiers
1. 658
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLE16K15"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI; 7 days post imbibition on water-agar. Mixed stage whole germinating seedlings from seed coat emergence up to two centimeters in seeds not showing obvious signs of germination were discarded."
BASE COUNT 208 a 126 c 124 g 200 t
ORIGIN

Query Match 7.8%; Score 161.6; DB 79; Length 658;
Best Local Similarity 58.0%; Pred. No. 4,6e+26;
Matches 385; Conservative 0; Mismatches 239; Indels 40; Gaps 4;

OY 248 gatggagaagaatacatatgctgagatgaaagcagagatataaagccttgagagatact 307
DB 2 GATTCTTAACCTTGCATGTCAGATAGACATGCGGGCTCCCAACATTCAGCGATCTC 61
OY 308 ctgcagcgaataatgctgcatacaaaaagcagcaaatcagcttgctggtgactccac 367
DB 62 ATTGTGAAGACTTACAGC-----AAGGACGTCCTGTCACCTCCCTGCTACACCTTT 115
OY 368 ctcttgacagggcgcccaaggtgagcggtgagttccatcccgagagcgagctactctgg 427
DB 116 CTGCTACCTTGCGACGTGAGTAGCGCGGCACTCCATATCCATCCGCGTTACTATGA 175
OY 428 attgagcagctacggtgctggaataattacttatttcaacggcgctatagcgagaa 487
DB 176 ATTCAACGACGACTGCTGACATATCTACTATTAATCTCAATGGCTATGAGATGAA 235
OY 488 at-----cgatcggggttcgagatctactctcccgagagactccagtg 535
DB 236 ATGAAGGTGTACTACCTCAATGATCCAAATTTGAGATATCCAAATTTGCCAAGGCTTCATTA 295
OY 536 ctggcccgagctgattaccgctcttcccttccctccacagagagatccgttca 595
DB 296 CTAAGAAAGCCAAAGATCTTCATCTTTTATTAGTTTCATCTACCTCAAAAGATGAGATA 355
OY 596 ctg-----atgaagagaaattgaaacttaagaagtgaaagaaacc 639
DB 356 TAGTTTGTCTTACCAACATTTCAAAAGCACTAGACACATTTAGATGCGAAGAAATCC 415
OY 640 taagctcttgatgacagcttgatgacgttgagccttgatgacgcacaaagccatgataa 699
DB 416 AAGGTACTTGTGAATCATTTATGATCATTTAGAGCTAGAGCCACATCAAAAGCATATTGAAA 475
OY 700 gtaagagatgatgcaatcgagccgctgattccctccgactcttgagcggtgaagatcc 759
DB 476 ATCAAAATTAATTTGGAATGAGACATTTGATCTTCATATCTTTGGGTGGAAGAAATTC 535
OY 760 ttcgagacagctcttcgagagatgcttcgagaaaggttcgaatgacagagatgctc 819
DB 536 ATTGGAATCTTCAATTTGTGTGATCTTTTTCAAA-----GTCAATGATGAGATCAT 589

OY 820 cgaatggttgacgaagaatccatccatcttcggtggttactcgttgcctggaagcttgc 879
DB 590 GGAATGCTTAACACAAAGCCTAATTCATTAATCTTATCTCATTTAGGAGCTATTT 649
OY 880 taat 883
DB 650 GAAT 653

RESULT 12
AI483541 476 bp mRNA EST 29-JUN-1999
LOCUS
DEFINITION EST249390 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
ACCESSION CLD25M13, mRNA sequence.
AI483541
VERSION AI483541.1 GI:4387465
KEYWORDS EST.
SOURCE
ORGANISM
tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 476)

REFERENCE
Alcala,J., Vrebalov,J., White,R., Materu,A.L., Viston,T.,
Holt,I.E., Liang,F., Upton,J., Romling,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Marlin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
Contact: David Fritsch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@CLEMSON.EDU.

FEATURES
source

Location/Qualifiers
1. 476
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLE25M13"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF"
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI; cLED - tomato carpel EST library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT 180 a 64 c 101 g 131 t
ORIGIN

Query Match 7.7%; Score 159.6; DB 40; Length 476;
Best Local Similarity 65.8%; Pred. No. 1,2e+25;
Matches 250; Conservative 0; Mismatches 124; Indels 6; Gaps 1;

OY 600 tgaagagaagaattgaaacttgaagaagtgaaagaaacctgaagcttctgtggaacagct 659
DB 57 TCAAGAGCAACTGAGACCATTAATGCTGACAAATAATCAAAAGCATCTTGCAATACAT 116
OY 660 ttgatgcttgagacgcgagatgagcctcaagccattgataagaagagatgataacgc 719
DB 117 TTGATGCAATTAAGCTGAGCCACTCAAACTTTTAAATAATTAATTTGAATTTG 176
OY 720 ggcgctgattacccctccgactcttgagacgtaaagatcccttcgagacaggtcttcgcgcg 779
DB 177 GACCAATGATTCCTTCATCATTTTGGGTGGAAGAAAGTTCATTTGAATCTTATTTGGTG 236
OY 780 gagattcttcgagaaaggttcgaatgacagagatgcttcgaatgcttcgagacgaatc 839

XR library construction kit Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-cold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erepding.*

BASE COUNT	196 a	100 c	162 g	160 t	1 others
ORIGIN					

Query Match	7.5%	Score 154.8	DB 72	Length 619
Best Local Similarity	57.1%	Pred. No. 1.5e-24		
Matches 342	Conservative 0	Mismatches 248	Indels 9	Gaps 3
Oy 776	ggcggagatttgcgcagaaagggctcgaaacgcagcttgcctcgaatgcttgaagcag	835		
Db 29	GACCAACATATATGGGCTGACACAGAAATTCACAGAGGATGAAATCATGCAATGATATGAT	88		
Oy 836	aatcctcgatcttcggttggttacgcttcgcttcggaagcttcgcttaacacagaaagtcg	895		
Db 89	AAGCCAAAAGGCGCTGTGTATATGATGATGCTTCGGGACATATAGCTACCTTTGGTGACGAG	148		
Oy 896	caatctgaagaagatagcaagaagggctgttagattcttggagagccgcttttgttggttgta	955		
Db 149	CAATATGGAAGATGCTACCTGTGTGTTGAAAGAGCTCTTGSGCTACTTTGTGTGGGTGTT	208		
Oy 956	agagtaaacgaagaaagaaagatgatgatgaattcgcagagagatctgaaacgaatggg	1015		
Db 209	AGAGCCCTGTAAGAGACTAAATCACTCTAAAGGTT---TTGAGAAAAAACAAGAAAGGC	265		
Oy 1016	aaatctgaccttggtgtgtctcgaattggaagtcgcagagccctccgtgtggatgttc	1075		
Db 266	TTGGTATGACATATGCTGCTCCACACTAAAGATCTTCTGCTCATGAGGCTATAGATGCTTT	325		
Oy 1076	gtcacacactgcgggttggaattcgactctagaagatatacttcggggttcgatgtg	1135		
Db 326	GTGACACATTTGTGGTTGGAAATTCACATTTAGAAATCTTATAGGAGTGTCGCATATTT	385		
Oy 1136	gcttccgcgaagtgtgttcgatcaagagacgaatctgaaagctgcatggaagatgtgtgag	1195		
Db 386	GCATACCATTTTGTGTCAGATCAAGCACAAATATGCAAAAGCTTATGCGCGAGATGTGGAAG	445		
Oy 1196	acgggtgtgagagctgaaagcctaattgagagaggttagcgctgttgaatgtgatyaatat	1255		
Db 446	ATCCGAATTTAGAG---CTCCTATTGACGACACACAGAGTTGTGCTCGAGAGCCTTAAG	502		
Oy 1256	agatgatcttaagaggttcttgatctggggagagaaaggttagaaacttagagagatgtc	1315		
Db 503	CACTCGCTAATAGGAGATATTCGA---GATGTAAAAATTCGAAGATGAAAAAGCTATGCC	559		
Oy 1316	ggcgaagtggaaagatttggcgaagaaagcattgtagaagaatgtgactcttcagttaaca	1374		
Db 560	ATGCGATCGAAGACCTGTGCTGTAAAGCTGCTACTGACGATGTAGCTCATATACAA	618		

RESULT 15

AM032414

LOCUS AM032414 579 bp mRNA EST 15-SEP-1999

DEFINITION EST275953 tomato callus, TMMU Lycopersicon esculentum cDNA clone

ACCESSION G1EC35B21, mRNA sequence.

VERSION AM032414

KEYWORDS AM032414.1 GI:5891170

SOURCE EST.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta

Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;

Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 579)

AUTHORS	TITLE	JOURNAL	COMMENT
Alcala, J., Vrevalov, J., White, R., Materin, A. L., Viston, T. Holt, I. E., Liang, F., Upton, J., Craven, M. B., Bowman, C. L., Romming, C. M., Fraser, C. M., Martin, G. B., Tanksley, S. D. and Giovannoni, J.	Generation of ESTs from tomato callus tissue	Unpublished (1999)	
Contact: David Fritsch			

FEATURES	Location/Qualifiers
source	1. .579

BASE COUNT	158 a	129 c	127 g	165 t
ORIGIN				

ORIGIN

```

Query Match      7.3%; Score 151.2; DB 63; Length 579;
Best Local Similarity 62.0%; Pred. No. 9,46-24;
Matches 294; Conservative 0; Mismatches 168; Indels 12; Gaps 3;

Oy 22 aaaaatgacagagcttcacgctcctcttggcacatctccagcacaggaacataaac 81
    |||||
Db 80 AAAAATGTGCAACCCCATATGCTCTATGGTGCAATTCCTCCACAAAGGTATATCAATCC 139
    |||||

Oy 82 cgccttcaatctgcacaagcgctctgcacaatgcgacattcaagtcacattctaacag 141
    |||||
Db 140 ATCTCTTCATTTGGCCAAAGGGCTATTCGAATGSGCATTTAGGTGACATTCACACATAG 199
    |||||

Oy 142 cgtctacgaatggcgcgcgaatgcacgaacgcgcgtggc---tcaaacgggctcatcaa 198
    |||||
Db 200 TGTTTTCCCCACGTCGTATGGCGCAAAATCCGGCTTCCACGGCACCCAAAGGCTTTAA 259
    |||||

Oy 199 ttgtgtgtcttccgcagggatgacagaggttaagc---cggagacgatggaa 255
    |||||
Db 260 CTGGCGGGCATTTCTGTGATGAGTTTGTATATGTTTCAAGCTTACGTGGATGATTTCTAA 319
    |||||

Oy 256 gaactacatctcggagatgaaagcagagataaaagccttgcgcgatactcttgcgc 315
    |||||
Db 320 ACGTTACATGTCCAGAGATTAAGAAAGTCGTGGGTCGCCAAACATTGAGGGATGTCATTTGAA 379
    |||||

Oy 316 caataatgtcatcaaaaaagcagcaaaatacagctctgtgtgtaccccaacctcttgc 375
    |||||
Db 380 GAGTTCAGACG-----AAGGACGTACTCTACGTCCTCTGTCTACACTCTTCTGCTACC 433
    |||||

Oy 376 atggcgcgacaaagtgcgcgttgaaattccattctccgagcgcgatactcttgattgagcc 435
    |||||
Db 434 TTGGGCGAGCTGAGAGTACGGCGGTGAACCTCATCTCCATCCACCCGTTACTATAGATTCAACC 493
    |||||

Oy 436 agctacaggtttgatataattacttaattcaagcgatacaggaacgaagaat 489
    |||||
Db 494 AGCAACGTGCTAGACATATCTACTACTATTACTTCAATGCGTATAGATGAAT 547
    |||||

```

Search completed: August 1, 2000, 14:38:28

Fri Aug 4 15:19:49 2000

us-09-147-955-5.rst

Page 12

Job time: 62700 sec

ALIGNMENTS

[illegible]

QY 901 ggaagatagacgaagagcgctgtagatctgtggaagcgcccttctgtggaagtagagct 960
 |||||
 Db 901 ggaagatagacgaagagcgctgtagatctgtggaagcgcccttctgtggaagtagagct 960
 |||||
 QY 961 aaacgaaggaaggaagagatgataagttgcaatggaagagttgaacgaagttgggaagt 1020
 |||||
 Db 961 aaacgaaggaaggaagagatgataagttgcaatggaagagttgaacgaagttgggaagt 1020
 |||||
 QY 1021 tgatcttggtgtctcattggaagtcctgacgcacatccctcgttggagagttctgag 1080
 |||||
 Db 1021 tgatcttggtgtgttcattggaagtcctgacgcacatccctcgttggagagttctgag 1080
 |||||
 QY 1081 aacgtcgggttggaattcgactctagaagatatacttctggaggttccggtgtgagcttt 1140
 |||||
 Db 1081 aacgtcgggttggaattcgactctagaagatatacttctggaggttccggtgtgagcttt 1140
 |||||
 QY 1141 tccgcaggtgttcgaatcgaaggaagcagtcgaacgtgataggagagttgtggaagcag 1200
 |||||
 Db 1141 tccgcaggtgttcgaatcgaaggaagcagtcgaacgtgataggagagttgtggaagcag 1200
 |||||
 QY 1201 tgtgaagagtgaaagcctaagaggaagtgacgtgtgtgagtgatgataagttgagagatg 1260
 |||||
 Db 1201 tgtgaagagtgaaagcctaagaggaagtgacgtgtgtgagtgatgataagttgagagatg 1260
 |||||
 QY 1261 tatgagaggttatgagatgagggagaaagagtagaagaaacttagagagagttgtgagaa 1320
 |||||
 Db 1261 tatgagaggttatgagatgagggagaaagagtagaagaaacttagagagagttgtgagaa 1320
 |||||
 QY 1321 gttgaagaggttggcgaagaaagcctaagaggaagatgagatcttcagtttaacaaactcaa 1380
 |||||
 Db 1321 gttgaagaggttggcgaagaaagcctaagaggaagatgagatcttcagtttaacaaactcaa 1380
 |||||
 QY 1381 ggtcttctctgagaggtgtgtagatctaaagaagtagaagttcccatagagcgaagaa 1440
 |||||
 Db 1381 ggtcttctctgagaggtgtgtagatctaaagaagtagaagttcccatagagcgaagaa 1440
 |||||
 QY 1441 ttcaaatctcctcgtatgagttgaaatcttctggcttaattcttctgagagttgtt 1500
 |||||
 Db 1441 ttcaaatctcctcgtatgagttgaaatcttctggcttaattcttctgagagttgtt 1500
 |||||
 QY 1501 ttcaaatctcctatgtaatttcaatctgagtgtaaatcttctgagtttcaatctgagaa 1560
 |||||
 Db 1501 ttcaaatctcctatgtaatttcaatctgagtgtaaatcttctgagtttcaatctgagaa 1560
 |||||
 QY 1561 aatataaatctatgtagagagctctcatcaaaatactcgaatcttcttcttctgagctg 1620
 |||||
 Db 1561 aatataaatctatgtagagagctctcatcaaaatactcgaatcttcttcttctgagctg 1620
 |||||
 QY 1621 gtcaaatctcctatcaatctggaanaataaattcgaatcttcccaatcttgaaccaa 1680
 |||||
 Db 1621 gtcaaatctcctatcaatctggaanaataaattcgaatcttcccaatcttgaaccaa 1680
 |||||
 QY 1681 gaaagaagatcaatcttgacccaataataaagaagatccaagtctctgtagaagtgct 1740
 |||||
 Db 1681 gaaagaagatcaatcttgacccaataataaagaagatccaagtctctgtagaagtgct 1740
 |||||
 QY 1741 gacgcagagctcctatcttccacgcgaattcttcaacgaagtttctgaattcttctgag 1800
 |||||
 Db 1741 gacgcagagctcctatcttccacgcgaattcttcaacgaagtttctgaattcttctgag 1800
 |||||
 QY 1801 caaaaatcggaactaacttctgacaaaatgaagttatataatgtaaatcttaaaaaaaca 1860
 |||||
 Db 1801 caaaaatcggaactaacttctgacaaaatgaagttatataatgtaaatcttaaaaaaaca 1860
 |||||
 QY 1861 ctcaagaatataaagccggaagtagtaaatatcctgacgaatcttgcgaatcttgcgc 1920
 |||||
 Db 1861 ctcaagaatataaagccggaagtagtaaatatcctgacgaatcttgcgaatcttgcgc 1920
 |||||
 QY 1921 tctatcttcaatcttctgtgtgttataataatcggttatcttcaatcttaataaata 1980
 |||||
 Db 1921 tctatcttcaatcttctgtgtgttataataatcggttatcttcaatcttaataaata 1980
 |||||

QY 1981 aaagttagatcatgatactgtgtgagatataatgagttatgtaatgtaacgatactt 2040
 |||||
 Db 1981 aaagttagatcatgatactgtgtgagatataatgagttatgtaatgtaacgatactt 2040
 |||||
 QY 2041 tct 2043
 |||
 Db 2041 tct 2043

RESULT 2

AB013596

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

Dh	2	AAATTTCCACAAAAGATGTCGCCGCCCGCTGCTAGAACGTTTCTGTGGCAAGGC	61
Oy	71	ccacataatcccgcccttcaattccgcgaagcgcttcgcaaatgcaaaatctaaatca	130
Oy	62	GCATTAATTCGGCCCTCCTCAATTCGGCAAGAACACTCTCTAAAGCCGGCACGTGAC	121
Oy	131	ttcttcacacgagctctacgaatgycgcgcacgcttcacgaacacgcgcgtgcttaacg	189
Dh	122	TTTTTCACGAGCGTTATGATGGCCCGCATAGCGCAACACAGCTCTCCGCCGTCCGA	181
Oy	190	-----gcatacaaatcttgctgctttcttcgcgcgagtgacgcagaggttacaagcc	241
Dh	182	AACCCACCGGGCTTGACTTCGTGGCGTTCTCCGACGGCTACACGACGGGCTGAAAGCC	241
Oy	242	ggaagcagatbggaagaaactacatgctcgagatgaaagcagaaggtataaagccctgagc	301
Dh	242	TGCGGCGACGGGAAGCGCTACATGCTCGAGATGAAGCCCGCGCTCCGAGCGCTTAAG	300
Oy	302	gatacctctgcagcaataatgctcgatcaaaaagcagaataatcagattctgctgctgag	361
Dh	301	-----AAACCTCTCTTCACAAACACACGACGTCAAGCTTCGTGCTTAC	343
Oy	362	tcaccactctctgcaatgycgcgcgaagctgycgcgtgagttcaatctccgcgaagcgtca	421
Dh	344	TCCACACTCTTTTGCACTGGCGCGCGGAGGTGGCGCGCTGAGTCCGAGTCCGACCGCCTT	403
Oy	422	ctctgagatgaaacgcagctcagctggttggatataatttacttcttaacacgcgtatgac	481
Dh	404	CTGTGGGTGAGACCGCCACCGTCTCTGTGATATTTACTTCTTACATTCGACGCTACGA	463
Oy	482	gagcaaatcagatbgcggtctgcatgcatctacgtccgcgaagacatcccaatgctgagc	541
Dh	464	GAGGAGATGTGACCGCGGTTCGACGAAATTC--GCTCCCTCGGCTTCCACCCCTGGAG	520
Oy	542	cagcgcgatttcagctctctctctctctctctccacgcagatgagaatctcgctcaactg	601
Dh	521	CACGCGAGCTTCGCACTTCTGCTGCTCCGAGACACCGGAGAAATTCGCGTTGATGATG	580
Oy	602	aagagagaatctgaaacttagaaggtgaaagaaaaacaaagctctgltgaaacagctt	661
Dh	581	AAGGAAACCTGGGAAACTTATAGACGTGAAAGAGGCGGAACGTTGGTGAACAGCTTT	640
Oy	662	gagcgcctgaaagcgcgattgcctcaagcgccatctgataagatcgcgaatgattcagc	721
Dh	641	GATGGTGTGACCCGAGTCACTCAAGGGCATTTGATAGTATGATTTGATCGGAGATCGGG	700
Oy	722	cagttgattcctctccgactctctggaacggttaagaatctctgcgaagctcttcgcgcga	781
Dh	701	CCGTGATATTCCTCCGCCCTTCTTGACAGCGGAGATCCCTCCGAAAGCTTTACGGCGC	760
Oy	782	gattctctcgaagaagggctcgaaatgcagacgagattgcctcgcgaatglttgcagaca	841
Dh	761	GATCTTTTGGAAAA--TCGAGAGAGAAATTACTCGTGGAGTGTGGACACGAAGCGG	817
Oy	842	cagttctggtgtgttaactgttctgcttcggaagacttcgttaatacgaagaagctcgaaatg	901
Dh	818	AAATCTCTCGGTGTATATGATGTGCTTTGGAGACGTTTGAAGCTTTCCAAAGGACAAATG	877
Oy	902	gaagagataagcaagaagggctgattagattctgtagaagagccgcttctgtgtgtgtaagata	961
Dh	878	GAGAGATTTGGGAAAGGGCTATTAGCTCGGAAAGCGCGTTTATATGATGATAGACGA	937
Oy	962	aa-----cgaagagaaaggggtatctgataaagtgcatagcagagaggttgaa	1006
Dh	938	CACAGCAATGACGACGGCGAAGAAAGAAAGAGATTTGAATTCATTTGGGGAATTTGAA	997
Oy	1007	cgaagtggaagaatctgatacttgctgctctcaaatcggaagctcgcaagctccctcgtg	1066
Dh	998	AAATATGGGGAATAATGTTGTGTGTGCTCGAGTTGGAGGTTCGGCGCACCTGTGTTG	1057
Oy	1067	gagtgcttcgtgaaacaactgcggtgtgaaatctgcactcaagagatatatcttcggtgt	1126
Dh	1058	GGATTTGTTCTGACCATTTGGTGGGTGAACCTGCGTGTGGAGACTTGAATTCGGGGT	1117

QY	1127	ccgaatgttgagcttccgcagctgctccgatacaaggaacgaatgagagctcaatgagagat	1186
Db	1118	ccgcctgtgtgcgcgtccgcacacagctgttgatcagacagacgaatggaaagctgattgagat	1177
QY	1187	gtgttggaagacgagtgctgtgaagaatgagactaaatgcagaggttagcgtcgttgatgtgat	1246
Db	1178	gccttgaggagacagggctgacagtgagaaagaaagaaagaggggt---ggggtttatgagact	1234
QY	1247	gaattcaggaagatctgatctgagaggtatctggaatgggaggaagaaagatcaggaactaga	1306
Db	1235	gagatagacaggtgtgtggaagatggtgatgattgattgggggtgagaaagacaaactagtaga	1294
QY	1307	gagagatgcgcgaagtgtgaaggaattcggcagaagaaactatagagaggaatgtgacttca	1366
Db	1295	gaaagagccatgaaatggaagacatttgcccaagagaagccatgaggagcattgactttca	1354
QY	1367	gttaacacccataagctctctctgataaggtctgtaagtatctaaagaacgaactaattgagt	1426
Db	1355	ctcagaagaattccaagcctttcttcatcaaatgttagcgtcttaattacacaaatagcgt	1414
QY	1427	ccc	1429
Db	1415	tcc	1417

RESULT

LOCUS	1458 bp	mRNA	PLN	20-MAR-1999
DEFINITION	AB013597	Perilla frutescens Pf3r6 mRNA for UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue, complete cds.		

ACCESSION
VERSION

VERSION AB013537.1 GI:411320
 KEYWORDS PF3R6; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.
 SOURCE *Perilla frutescens* (variety:crispa, cultivar:Shikun) leaf cDNA to
 mRNA.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asterales; Gentianales; Lamiales; Lamiales; Perillales.

AUTHORS

TITLE Kusumi, T. and Saito, K.
Molecular cloning and biochemical characterization of a novel
anthocyanin 5-O-glucosyltransferase by mRNA differential display
for plant forms regarding anthocyanin

JOURNAL

MEDLINE
 9916/509
 2 (bases 1 to 1458)
 yamazaki,M., Saito,K. and Gong,Z.

JOURNAL

Yamaaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. & Biotech., Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan
(E-mail: yamaaki@p.sc.chiba-u.ac.jp, Tel.: +81-43-290-2905, Fax: +81-43-290-2905)

FEATURES

```

source
1. .1358
/organism="Perilla frutescens"
/cultivar="Shikun"
/variety="crispa"
/db_xref="taxon:48386"
/tissue_type="leaf"
29. .1360
gene
29. .1360
/gene="PF3R6"
CDS
29. .1360

```

```

/codon_start=1
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase
homologue"
/protein_id="BAA36422.1"
/db_xref="GI:415561"
/translation="MRRRYLATFPAQGHINPALQFAKRLKACTDVTFTSYAMR

```


COMMENT

E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/th1/>
this fragment has an overlap with ATCHRIY37 at the 5' end and an
overlap with ATCHRIY39 at the 3' end.

FEATURES

source	Location/Qualifiers
gene	1..198944 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 4845..5598 /gene="AT4g14060" join(4845..5031,5330..5598) /gene="AT4g14060" /note="strong similarity to major latex protein type 1, Arabidopsis thaliana; contains EST gb:A1994471.1, T43693, A1099764, H36765, T43905, A1998305.1, R29829, T22625, R90142, T41926, T22170, W43041, R65053, AA712594, AA712597"
CDS	/codon_start=1 /product="major latex protein like" /protein_id="CA878448.1" /db_xref="GI:7268111" /translation="MATSGTYTEVPLKGSAAKHKRMENHLPDPAIGHHIGVTV HDGEMDSIGALIKWNYTLGDKPEMFEKREIDEENAAVTFCVLECHVEQLKVVDTIL OFIOKSPDIIYCKITMTWEKRADSDPEPSNYMKLVKSLAADDAHVLA"
exon	4845..5031 /gene="AT4g14060" /number=1 5032..5329 /gene="AT4g14060" /number=1 5330..5598 /gene="AT4g14060" /number=2 complement(9562..9654) /number=1 9562..15425 /gene="AT4g14070" complement(join(9562..9654,9750..9807,9883..9998, 11097..10176,10276..10435,10594..10731,11065..11145, 11247..11300,11383..11541,11653..11758,11882..11949, 12156..12305,12439..12554,12626..12834,13288..13360, 13958..14072,14674..14859,15291..15425)) /note="strong similarity to AMP-binding protein - Brassica napus; contains putative AMP-binding domain signature AA235-266; contains EST gb:A1999282.1"
exon	/codon_start=1 /product="AMP-binding protein" /protein_id="CA878449.1" /db_xref="GI:7268112"
exon	/translation="MASTSLGSPSTLLSGSPROPDPDRPILSGHESVRIIPRRP REKVKRSFPLESSFSGDALRSSEKRAVPDIRSSAKTGDYVALVDYDRPPEK LITKOLREIIDFAGSLVGLVKADEKIALFDNSCNLWSDGAVNVKSRSSVEE LQIVRIHSEVAIVYDNPENRIAESFSPKSLRFLILMGKSSLVTOQMQLPVSS YAEIINOGQESRAKLSANDTFSYRNQFTDSDTAALIMYTSGTGNPGVMILHNL HOIKHLSYVAQAGDKFLSMLPSMAHERASFLFTGCVQMYTSGIRYKLDIKRY OPNYIVSVPLVETLYSGIQKISASAGKRFATLQIKVMAWEMKRIEKGCLTK EOKPEPMIVAVDMLARVIALIMPLMLAKKLIYKIIHSSIGISKAGISGGSLPI HDKPFEAIGVILONGYGLFETSPVCARTLSCNVLGSHGPHMGTEPKIYDPENNY LPBSKGIYRGQVAKGYKYPSTTKOVLNCSGFMNCGMATAPIHSGRSRHC GVYVLEGHARDIYLVSTGENVERPLEIEBAKMSRVLEQIVIGDRRRLGAILIPKE EAQRVDEPSEKELKSLVLOELKRMVSECSFOVGPLVLDPPFTIDGMLPTMKIR DMVAKYKEEIDQYS"
intron	complement(9655..9749) /number=1
exon	complement(9750..9807) /number=2 complement(9808..9882) /number=2 complement(9883..9998)

intron	/number=3 complement(9999..10096) /number=3 complement(10097..10176) /number=4 complement(10177..10275) /number=4 complement(10276..10435) /number=5 complement(10436..10593) /number=5 complement(10594..10731) /number=6 complement(10732..11064) /number=6 complement(11065..11145) /number=7 complement(11146..11246) /number=7 complement(11247..11300) /number=8 complement(11301..11382) /number=8 complement(11383..11541) /number=9 complement(11542..11652) /number=9 complement(11653..11758) /number=10 complement(11759..11881) /number=10 complement(11882..11949) /number=11 complement(11950..12155) /number=11 complement(12156..12305) /number=12 complement(12306..12438) /number=12 complement(12439..12554) /number=13 complement(12555..12625) /number=13 complement(12626..12834) /number=14 complement(12835..13287) /number=14 complement(13288..13360) /number=15 complement(13361..13957) /number=15 complement(13958..14072) /number=16 complement(14073..14673) /number=16 complement(14674..14859) /number=17 complement(14860..15290) /number=17 complement(15291..15425) /number=18 complement(16137..16174) /number=1 16137..17732 /gene="AT4g14080" complement(join(16137..16174,16252..17604,17687..17732)) /note="Che CDNA A2668 is cognate to this gene; gene A6 for anthr-specific protein - beta-1, 3-glucanase homolog; contains glycosyl hydrolases family 17 signature AA274-287; prokaryotic membrane lipoprotein lipid attachment site A48-18; contains EST gb:AM004339.1, 218400"
intron	/codon_start=1

/product="A6 Anthr-specific protein"
/protein_id="CAB78450.1"
/db_xref="GI:7268113"
translation="MSLAFLEFLITLVFSSSCGATRFQGRVYMQRKMLDLAKIGI
NYGRGNLSPYOSINFTRKAGHVKLYPADESJTLISQNLVYITVPHQHTA
LSSNOJLADENWNTILPYQTOIRFLVNEILTSNSGVNLYPAIRKLVNLR
LHGJLHKKYGPFLADSLRSPSPNGFREIEIPVMLPLFLNGTNSEFLNHP
YFMSRNPMTSLDEALFOGHSTYDPTQGLVYNLDLDLDFEAMRKLQPHRL
AISEGMPNPDIDEGANILNATVNNLLKKMSAPRIGTPSRGGLPPEVSLF
NEMOKSGCTORHKGLHPDGPRIYDNPCTGRTFNPRLPPIINVPYKQGVNCP
VEANTELLEPRLRMACQSNITCALAPGECIEPVSITMHASYALNSYAWDFRMS
IQCFTNGLAHETINFGNDRKFRSVTL"
intron
complement(16175..16251)
/number=1
exon
complement(16252..17604)
/number=2
intron
complement(17605..17686)
/number=2
exon
complement(17687..17732)
/number=3
exon
18189..18494
/gene="AT4914090"
/number=1
gene
18189..25761
/gene="AT4914090"
join(18189..18494,18694..18787,18938..19018,19571..19724,
19772..19833,21488..21524,22014..22792,22844..22948,
23060..23200,23280..23566,24375..24631,24686..25140,
25192..25311,25445..25614,25665..25761)
/gene="AT4914090"
/note="similarity to proton-coupled peptide transporter
PEPT, Rattus norvegicus; Contains Prokaryotic membrane
lipoprotein lipid attachment site A453-545;Prokaryotic
membrane lipoprotein lipid attachment site A4955-965;
contains EST gb:A1998193.1"

Query Match 15.1%; Score 310.6; DB 8; Length 198944;
Best Local Similarity 58.2%; Pred. No. 3.3e-52;
Matches 805; Conservative 0; Mismatches 509; Indels 69; Gaps 12;

36 ctccagctctcttggccacatccacgacagagacacataatccgccttcacatcg 95
DB 21210 CACATTACTTCTCTTAACATTCCACGCCAGGTACATCAACCGCGCTCACTAG 21151
96 ccaaggctctcgcaaatgcccgaatcaagtcacatcttcacagcgctcagatgac 155
DB 21150 CCACCGCGCTCATCCACACGAGTCAACGCTCAGATTCACCGCAGTCTCTGTCACC 21091
156 gccgagctcagaaaccgcgcctcgctcaaaacgggtcacaattctgtcgtttccg 215
DB 21090 GACGATAG---CGGAGCACCACTTCACAAAAGGTTATCC---TTCCGTTGGTTCAACCG 21037
216 accggtatgacgaaggttacagcccggagacgaagggaagaacatactgcgaagatga 275
DB 21036 ATGGATTTCGACGAGGCTCTCAAGTCATTGGAAGACCGAAGAAATTCATCTCCAGAACCA 20977
276 aaagcagaaggtataaagccttgagcatalactcttgacgacataatctgcataaaa--- 332
DB 20976 AACGATGTGGTTCAAGCGCCCTGAGACATCATCAAGCCAAATTTCACGCCACACCG 20917
333 aaagcagaanaatacgtctcgaggctacccctcttgcatggcgagcgaagtgag 392
DB 20916 AAACGAGACCTATCACCGCGGTAATCTACTCTGTTCTCCCGGGGTTTCTACGAGTAG 20857
393 cgcgtgagttcaccatccggagcgctactcttgatgtagagcaactcaggtgttgatga 452
DB 20856 CGCGTAGATTACACCTCCCACTACACTTCTCTGAGTTTGAACACACTACTACTAGACA 20797
453 tacttaacttattcaacgactaagcagaanaatcgaatcgagggttcgagatcattc 512
DB 20796 TCTACTACTACTACTTCAACACCTCTTACAAACATCTCTTCGACGCTT---GAACCGATTA 20740
513 actgagccggagagctccagtgctgcccagcgtagattaccgcttctcttctctc 572

DB 20739 AATTATGC---GAACCTGCCACTGATCACACCGGTACCTCCGCGTGTCTTCAACCTT 20683
573 ccaagcatalgagatctccgttcactga---tgaaaggaanaattggaacttgaaggtg 629
DB 20682 CGAAGGCAATTACCGCTGACGCTTGTAAGTACTTAAGAACATATCAACGCTCGAAGCG 20623
630 aagaaaaaacctaaaggtcttgtagaagaagcttgatgacgtctggaagctatagcgtc 689
DB 20622 AATCAAAACCTTAATATTTCTGTTAACAATTTCTCTCTTTGGAACACATGCTTTAACT 20563
690 ccatgataagtaagatgagatgcatcgaaacggcgcttgatcttcctccgcatcttgagc 749
DB 20562 CTGTTGAGAAACTCAAGATGATCCCAATCGAGACCGTGTCTTC-----CTCCGAGG 20509
750 gtaagatccctcggagacaggtcttcggcgagatttgctcgagaagggcgcgaatgacg 809
DB 20508 GTAAACCGATCTTTTCAAAATCTTCC-----GACG 20479
810 acgaltgcctcgaatggttgagcagaatccctcgatctcgttggttgatcttcgttcg 869
DB 20478 AGGATTACACGAATGCTTAGACTCGAAGCTCGAGATCACTGATTACATTTCTCTTAG 20419
870 g---aagcttgtaataagcagaagtcgaaatggaagaagtagcaagggcgctgttag 926
DB 20418 GCACACACGCCGATGATTTACACAGAACACATGGAACCGCTTCTCACGCGCTGTAG 20359
927 atgtgagagcgcttcttggtggtgtaagagtaacga-----aggagaagaggt 978
DB 20358 CTACAAACGACCGCTTTTATGATCGTAGGAGGAGAAAATCCAGAAAGAAAGAA 20299
979 attgaaagtgacatgagagagtggaagagtgaggga---aatttactcttggtgtctc 1037
DB 20298 ATCGTTTCTTGAATTATATAGAGAGATGATCGAGATGAGTGTGATGGATGGTGTCTC 20239
1038 aattggaagctcgtgacgacatccctcgttggaatcttcgtgacacacgcggttggaatt 1097
DB 20238 AGACAGCGTTTTCGCCCATATTCGTGCGATGTTTGTGACTCATTTGTGTGGAATT 20179
1098 cgaactgagagatatactcttcgggttcgagatggttggtcttccttcgagatggttcgattc 1157
DB 20178 CGACGTTGAGAGATTAGAGAGTGTGAGAGTGTGCGGTGCTTCGACGATTTGCTGATTC 20119
1158 aagggaagatgacgaacgcgacgagagatggttggaagagcggtgtggaagtgaagacta 1217
DB 20118 AGTGTACAAACGCGCAACCTGTGTGAGATACGTGGAGGATTTGAGATGAAGGTAAGG--- 20060
1218 atgaagaggtgagcgtcgttcgtatgtagatgaatgagagatgatacttgagaggttaatg 1277
DB 20061 TTGGGAGAGAGAGATGTGATGGGAGAGATTGAGAGGTGTTGAGAAAGTGATGA 20002
1278 atggggagagaagaagatgaagaaacttagaagaggtctcgcaatggaagatctgacaa 1337
DB 20001 GTGGTGAGAGAGAGCGGAGAGATGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAG 19942
1338 gaaagcattggaagagatcgatcttcagttacaaactcaagctcaagcttctctatgag 1397
DB 19941 TTGATGCCCGGACGCGAAGGTGAGACCTCGGATTTGAATCTTAAAGCTTTGTGACACAGG 19882
1398 ttg 1400
DB 19881 ATG 19879

RESULT 6
LOCUS APTCAO 200576 bp DNA PLN 28-JUN-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
ACCESSION 297335
VERSION 297335.2 GI:5280985
KEYWORDS
SOURCE thale cress.

ORGANISM	Arabidopsis thaliana	gene	5905..7913
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	CDS	/gene="d13005c" complement(join(5905..6149,6243..6344,6436..6826,7236..7913)) note="strong similarity to glycine hydroxymethyltransferase (EC 2.1.2.1); contains EST gb:1565775"
REFERENCE	1 (bases 1 to 200576)		
AUTHORS	Bevan M., Slinkema W., Murphy G., Wambutt R., Pohl T., Terry N., Kreis M., Kavanagh T., Entian K.D., Rieger M., James R., Puidomenech P., Hatzopoulos P., Obermayer B., Duesterhoft A., Jones J., Palme K., Anstorge W., Delsen M., Hancock I., Mewes H.W., Scheller C. and Chaltatzis N.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 200576)		
AUTHORS	EU Arabidopsis sequencing project.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@hsrc.ac.uk		
COMMENT	On Jun 30, 1999 this sequence version replaced gi:2244747. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with AF061 at the 3' end.		
FEATURES	Location/Qualifiers		
source	1..200576 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 1..63289 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="BAC T4M020K8" 1753..4428 /gene="d13000w" join(1753..3194,3261..4428) /note="similarity to Cf-2.2, Solanum pimpinellifolium" /codon_start=1 /product="disease resistance Cf-2 like protein" /protein_id="CAB10171.1" /db_xref="GI:2244748" /db_xref="SPTRMBL:023253" /translation="MTITWISGICITGICSTLIYAKDICTPDQDALLEFKNESPSPSDAMLIOTTKMRNNTDCCSGGICSDPRGTGVVVELDIGNSDLMKRLNSLFRRLQHLQSDLSYNDLSCTLPDSSGNERVYLVNLGSCNLFGBIPISLNSLYLVDLSDYNDLGEILDSMGNLKLRLVSLSCFTGIPISLNLTYLVDLSDYVYFGEIPDSMGNLSLVLNLRHCNFFGKIPISLGLSNLTDIDISGNSFSGTIPISLNLTDFOMLNLSTLNVDLNSOFRKAMPMSMSLKLDAFDISGNSFSGTIPISLNLPSLTKDGTNDPFGPLKIGNISNLOELTYGNNINGPIPRSLIKVGLASLISFMDGIVDPSITLQKLSRLDLSGNNISSHLPSHMHLLISCNISQPKLEMOGSLYHDIISANOIGQVPEMLRHPITSFASNKSGEIPRAYCEIGTVLSNNFSGSIPCEISIKTKLILHLKNNISGVTIPESLHGLNLSLDVGSNRLSGOPEKSLINCSYLOFLVNEENRINDTFPSWLKSLPNQLLVNFSNFFGPIFSPDLSLFSKLRFDISENRSGLVPSDFVGSVMSFVDIIDNTPGTVVGDQESFHSVYLTIGLNLMEVSGSEIYKTIIDVSGNRLGEDIPESIGLILVLNMSNNAFTGHIPELSLNLNLOSILSONRSLSGIPGELGELTFELARNNSYELGPIPOGTIOSONSSSFAENPLCGAPLKKGGGEEDKEKEKDGSLGWSVAAAGVVPPLFGCLAIGHILSYKNDWMRIFSCFSSPLKK"		
exon	1753..3194 /gene="d13000w" /number=1	exon	5905..7913 /gene="d13005c" complement(join(5905..6149,6243..6344,6436..6826,7236..7913)) note="strong similarity to glycine hydroxymethyltransferase (EC 2.1.2.1); contains EST gb:1565775"
intron	3195..3260 /gene="d13000w" /number=1	intron	6150..6242 complement(6150..6242) /number=1
exon	3261..4428 /gene="d13000w" /number=2 complement(5905..6149) /number=1	exon	6243..6344 complement(6243..6344) /number=2
		intron	6345..6435 complement(6345..6435) /number=2
		intron	6436..6826 complement(6436..6826) /number=3
		intron	6827..7235 complement(6827..7235) /number=3
		intron	7236..7913 complement(7236..7913) /number=4
gene	1753..4428 /gene="d13000w" join(1753..3194,3261..4428) /note="strong similarity to adenosylhomocysteinase (EC 3.3.1.1)" /codon_start=1 /product="adenosylhomocysteinase" /protein_id="CAB10173.1" /db_xref="GI:2244750" /db_xref="SWISS-PROT:023255" /translation="MALVEKTSRGREYVYKQKSOADFGRLLELAEVEPMGLMACRT EGPSPQPRKARITGSLHMTIOATVLETLTALGAEVAKSCMISTODPAAALAAARD SAAVYAMNGETLQERYWCKTEPAAEDMGSGGPRDIYVDDGDAITLIDEGKAEITEKT GGVPEPTSDPERFOIVLSIIKESLQVDPKATKMKERLVGSEETTVKRLTMOOQ NOTLLEPAIVNDSVTSKSEPNLYGCNHSLPDELMAKTVMTAGKAAVTCGVDGKG CAAMAKKTVAGAVYTEIDPICALQALAMEGLQVLEDDVYSEADIFVYTTGNDDIYWD HRRKKRNAAIVCANIGFDNEIDMGLGETYGVAVRITIKPOTBRWVPEPKADIIYLAEGILNMGATGHPGVPVMSGSTNNOYVIOQELMNERKSGVKKVYVLPRLHDEKVALL HIGKLGATLTSLKDQSDYVSIPIEGYKRPPIHRY"	gene	12823..14568 /gene="d13010w" join(12823..13533,13822..14568) /note="strong similarity to adenosylhomocysteinase (EC 3.3.1.1)" /codon_start=1 /product="adenosylhomocysteinase" /protein_id="CAB10174.1" /db_xref="GI:2244751"
CDS	1753..3194 /gene="d13000w" /number=1 complement(16093..16270) /number=1 16093..16501 /gene="d13015c" complement(join(16093..16270,16317..16501)) /note="weak similarity to Walleys dermal sarcoma virus gag polyprotein" /codon_start=1 /product="hypothetical protein" /protein_id="CAB10174.1" /db_xref="GI:2244751"	CDS	12823..13533 /gene="d13010w" /number=1 complement(12823..13533) /number=2 complement(13533..13822) /number=3 complement(13822..14568) /number=4 complement(14568..15657) /number=5 complement(15657..16093) /number=6 complement(16093..16270) /number=7 complement(16270..16317) /number=8 complement(16317..16501) /number=9 complement(16501..16931) /number=10 complement(16931..1753)

Db 80333 ATG 80331

RESULT 7
LOCUS AB000623 1594 bp mRNA PN 05-FEB-1999
DEFINITION Nicotiana tabacum mRNA for glucosyl transferase, complete cds.
ACCESSION AB000623
VERSION AB000623.1 GI:1805358
KEYWORDS JIGT: glucosyl transferase.
SOURCE Nicotiana tabacum (Strain:BY-2) suspension culture cDNA to mRNA.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Asteridae; Solanaceae; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1594)
AUTHORS Kojima,H.
JOURNAL Direct Submission
TITLE Submitted (23-JAN-1997) to the DDBJ/EMBL/Genbank databases. Hisae
Kojima, school of agriculture,nagoya university, laboratory of
biochemistry: nagoya, furo-cho, chikusa 464-01, Japan
(E-mail: hisaek@nagrl.agr.nagoya-u.ac.jp, Tel:052-789-4098,
Fax:052-789-4094)
2 (bases 1 to 1594)
AUTHORS Kojima,H., Hashizume,K., Imanishi,S. and Nakamura,K.
JOURNAL jasmonate-induced potential glucosyltransferase from tobacco
TITLE suspension cell
FEATURES
Journal Unpublished (1997)
source Location/Qualifiers
1..1594
/organism="Nicotiana tabacum"
/strain="BY-2"
/db_xref="taxon:4097"
/rname="BY-2"
/rsize_type="suspension culture"
52..1455
/gene="JIGT"
52..1455
/gene="JIGT"
52..1455
/standard_name="jasmonate-induced potential
glucosyltransferase"
/codon_start=1
/product="glucosyl transferase"
/protein_id="BA19155.1"
/db_xref="GI:1805359"
/translation="MENIKNECHVLAIPFGGHINPSLOFSKILINLGVYTLSSSL
SAFNRIKMLPRTEGLTAPSPSGDYGNGKSPDDYHFNPAIKSGSEPIANLITSKA
KNGYPTPTVITYIIMDNAGSAKRLHPSLPMIOPATVVDIYYRTFNFYRKNYD
SDDQITLPLGLSSDSDFSEFVDVSKDMVESIKRQIEITLNEENPRYLIVTFD
AELNALRVLNVTWVGILPSPSFLDERDNDKRDNFPAADMESENNYEMWLDARANK
SVIYIAFGSVAIEISSOMNEIISQILKCGRPLWVIREITLNGEKPEKXICDELEKI
GRIVMSQMEVLEKHSYVGCFLHCGNNLSLESIASGVPITACPIPNDDICNAKLIOD
VYKIGRVANKKEGIIRKDEFQKCIETVMGDAEGBELKRNMAKMDLAKESTKENSS
SNVNLKAYVNECLIGHC"
polyA_site 1594
BASE COUNT 514 a 238 c 325 g 517 t
ORIGIN
Query Match 14.0%; Score 288.6; DB 7; Length 1594;
Best Local Similarity 54.1%; Pred. No. 7.8e-48;
Matches 766; Conservative 0; Mismatches 609; Indels 42; Gaps 7;

Db 165 CACTTATCTTACGCTATGACCTTTCATGCGATA--AAATCTCCCAAAATAGA 221
Oy 187 cgggctcatcaatcttgctgcttccgacgggcatgaagcgggttaccgcccggaga 246
Db 222 AGGATTAACTTTGCTCCTTTCTCTGATGGCTATGATGAAATTTCAAGTTCAATTGA 281
Oy 247 cgaatggaagaactacatcgtcgagatgaagaacgaaggtataaagccttgagatc 306
Db 282 TGATTACCATTTATTCATAATTCAGCATTAATAATCTCATGTTCTGTAATTTATGCGCAATTT 341
Oy 307 tcttcgacgaataatgctgcatgataaaagaacgaagaatcgcgttcgagtgatccca 366
Db 342 AATTAAATCCAA-----GGCAAAAATAGTTACCCCTTTACGCTGATTATTAACATAT 395
Oy 367 cctcttgatcagcgcgccgaaggtgcgcgatgagtcacatcgcgagacgcgcatcgt 426
Db 396 TCTTATGATGGGCTGGTCACTACGACCAAAAATTAATACATACCTTCTCAATTAATTTTG 455
Oy 427 gattgagcagcagcagcgtgttgatatacttacttacttca-----acgctatag 480
Db 456 GATTCAACACAGCTACAGTTTGTATTTATTTATGATGTAAGCTTATGATGCGAGTTGA 515
Oy 481 cgaagaatacgaatcgcggttcggaagctatcaactcccgagaggaactccagtgctgac 540
Db 516 TTTCAAGAAATTAATGATTTCTCAGATCAAAATTAATGAGCTACACAGCTTGCCTTCATAG 575
Oy 541 ccaagctgattacgcgtcttcccttccctcct-----tccacgaatgagagatcgttc 594
Db 576 TTCAAGGATTTTCTCTCATTTTGTGTTGATGATGTAAGCTTATGATGCGAGTTGA 635
Oy 595 actgataagaggaatctggaactcttagaaggtgaagaagaacgaagtcctgtgtgaa 654
Db 636 GTCAATTAAGGCAAAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 695
Oy 655 cagcttgatcgtctggaagcctgacgctcgcacgaagcattgataagtcgaagatgctg 714
Db 696 TACTTTGATGATGATGGAAGCTTAATGCTTTGAGGGTTTGAATAATGTAAGTATGATGTTG 755
Oy 715 aatcgcgcgctgattccttccgcatcttcttgagcggtaagaatccctcgacaggtctt 774
Db 756 AATTGACCTTAATTCCTTCATCTTCATGATGAAAGATGCAAAAGTAATTTT 815
Oy 775 cggcggaagattctcgagagaagggcgaatgacgaagatgctcgaatgctgagcac 834
Db 816 CGCGGCGATATGATGCAAG-----TGAATAATATTAATCATGAGAGTGCTTGATGATG 866
Oy 835 gaatcctcgatctcggttggttgaagtcgttcggtgaagccttgatatacgaagaatc 894
Db 867 AAGGCGAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 926
Oy 895 gcaaatggaagatcgaagaagggcgttagatcgtcgagggccggtttctgtggtg 954
Db 927 ATGATGGAAGAGATTAGTCAAGAGATTGTGAATGTGGAAGACCATTTTGTGGGTGAT 986
Oy 955 aagaagtaaac-----gaagagaagaagatattgataagttgcatggaaggttgaa 1005
Db 987 TAGAGAAACCTTAATGCGGAAAMACCGAGGAAAGATTGATGTAAGTGAATTTGA 1046
Oy 1006 acgagtgaggaaatgtatcttgctgctcgaatctggaagtcctgacgcatccctcgt 1065
Db 1047 AAGATATAGGAGATATAGTGTGTGTCGTCGCAAAATGGAAGTCCCAAAACGCTGTCGT 1106
Oy 1066 ggaatgttctgagcacacgcgcggttggaatcgcgactagagagatatacttcggtg 1125
Db 1107 AGGCTGTTTCTTGACTCATATTTGATGGAATTCGACTGGAACCTTAAGCTTGTGTTGT 1166
Oy 1126 tccgagtgtgcttcccgagagtgcttcgaataaaggaagcgaatgcaagcgtgataagaga 1185
Db 1167 GCCAATGCTGGCATGCCGATTTGGAATGATCAAAATTTGCAAGGAAAGCTTCATTCAAGA 1226
Oy 1186 tctgtggaagacgggtgtgagagtgagtcgaataaggaaggttagcgctgtgagtgta 1245
Db 1227 TGTTTGGAAGATTGGTGTGAGTGAAGCTAATTAAGAAGAGTA--TCATTAAAGAGA 1283

```
OY 1246 tgaattcaggagatctatcgaaggctatgatgggaggaagaagtaggaacttag 1305
DB 1284 CGAGTCCAGAGGCTGTATGATGATCGTAATGAGGAGATGCTGAGGAGAGAGCAATTGAG 1343
OY 1306 agagatgctgcagcaggaaggatttgcgaagaagaacatcagcgagatgcttc 1365
DB 1344 AAGAAATGCCAAATGAAAGATTTGGCTAGGAATCTAGCAAGAAATAGTTTCATC 1403
OY 1366 agtcaacnctcaggtcttctctgatcgaggttga 1402
DB 1404 TAAATGTAATCTTAAGGCTATGTTATGTAATGTTTA 1440

RESULT 8
ATCHRIV42/c
LOCUS ATCHRIV42 195068 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment NO. 42.
ACCESSION AL161542
VERSION AL161542.2 GI:7268303
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana
            thale cress.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsi.
            1 (bases 1 to 195068)
REFERENCE 1
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            loeckemips.biochem.mpg.de,mayeremips.biochem.mpg.de,project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV41 at the 5' end and an
            overlap with ATCHRIV43 at the 3' end.
FEATURES
            location/Qualifiers
            1..195068
            /organism="Arabidopsis thaliana"
            /variety="Columbia"
            /db_xref="taxon:3702"
            /chromosome="4"
            9917..10066
            /gene="At4g15560"
            /number=1
            join(9917..10066,10350..10475,10559..10874,10981..11252,
            11340..11613,11691..11998,12081..12362,12441..12725,
            12813..12953)
            /gene="At4g15560"
            /note="GT/AC sites ignored to preserve continuity; light
            inducible gene required for chloroplast development;
            Contains prokaryotic membrane lipoprotein lipid attachment
            site AA452-462; Prokaryotic membrane lipoprotein lipid
            attachment site AA591-601; Transketolase signatures
            AA102-121; Transketolase signatures AA505-521; contains EST
            gb:AM004332.1, A1997953.1, W43562, N96015, W43388"
            /codon_start=1
            /product="DEP (CIA1) protein"
            /protein_id="CA878598.1"
            /db_xref="GI:7268304"
            /translation="MASSAFAPSYIIRKGLSTDSCKSTSLSSRSRLVMDLSPCK
            PNNNSHNRAYVCAISLAKEGYENRPPPLDPTINPIIMKRLSVKELKQASDLR
            SDIIVNSKRTGGLGSSGLVVELTYALHYIRNTPODKLIMDVGHQSPHKLILGRBK
            MPTWRQNGLSGFTKRGESEHDCFTGHSSTFTISAGLQMAVGRDLKKNNNVAVIGD
            GAATAGQAV EANNMAGYLDSDMIVILNNKQVSLPTATLDGPPPVGLSALSASRLQS
            NPALRELRVAGKMTKOIGGPMHQLARVDEYAKMISCTGSSLFELGLYIIGPVG
            HNIDDLVALLEKSTRTPGLVLIHVTEKRGYPAERADKYGAGVVKFDPATGRQF
            KTNKTSQSYTTYFAELVAEAEVDKQVVAIHAAAGGTGTLNLFQRRRPTRCFVGIAE
            OHAVTPAAGACAGLPCFCAIYSSFMQRAYDOYVHDVDLQKLPYRFAMDRAGLYGADG
            PTHRGADPTVEMACLPNMTVMAPSDADLFNNVATAVATIDRPSCFNTPKNGIGVAL
            PPGKRGVPEIGKRIILKEGRVALLGSGSAVOSCTGAVALBERGLNVAVARFCK
            PLRALRIIRLSLAKSHEVLIIVDEGSIIGGRSHVDFLALDGLDKLKRPMVILPDRIY
            DGGAPDQDLAEAGIMPSVIAATVALNLIGAPREALF"
            9917..12953
            /gene="At4g15560"
            10067..10349
            /gene="At4g15560"
            /number=1
            10350..10475
            /gene="At4g15560"
            /number=2
            10476..10558
            /gene="At4g15560"
            /number=2
            10559..10874
            /gene="At4g15560"
            /number=3
            10875..10980
            /gene="At4g15560"
            /number=3
            10981..11252
            /gene="At4g15560"
            /number=4
            11253..11339
            /gene="At4g15560"
            /number=4
            11340..11613
            /gene="At4g15560"
            /number=5
            11614..11690
            /gene="At4g15560"
            /number=5
            11691..11998
            /gene="At4g15560"
            /number=6
            11999..12080
            /gene="At4g15560"
            /number=6
            12081..12362
            /gene="At4g15560"
            /number=7
            12363..12440
            /gene="At4g15560"
            /number=7
            12441..12725
            /gene="At4g15560"
            /number=8
            12726..12812
            /gene="At4g15560"
            /number=8
            12813..12953
            /gene="At4g15560"
            /number=9
            14413..14436
            /gene="At4g15570"
            20338..24557
            /gene="At4g15570"
            20338..20373
            /gene="At4g15570"
            /number=1
            join(<20338..20373,20767..20852,20962..21012,21256..21367,
            21467..21572,21653..21750,21989..22117,22182..22274,
            22625..22678,22957..23015,23122..23206,23286..23765,
            24183..24269,24366..24557)
            /gene="At4g15570"
            /note="Similarity to SEN1 protein, yeast; Contains
            ATP/GMP-binding site motif A (P-loop) AA13-20; contains
            EST gb:N65040"
            /codon_start=1
            /product="SEN1 like protein"
            /protein_id="CA878599.1"
            /db_xref="GI:7268305"
```

/translation="VGLSRKSFVLIQCPGCTGTOTLISLCAIMHATPARVOSKTD
HEKRGIOHTIOEKYINWGRASPILISVNFPRDAIMPEDDGDFPTSGNELPEVNA
SKRYRLVAVCAPSNSALDEIVLRLLSSGLRDENAOTYPRIVRIGLKAHHSVASVL
DLVAKRGDVSALDKPKQSTGTIDISIRTAILEEAAIWMGFNRIVAFVATLSFSSAL
LAKSNRGFDVAIDEAQAVPATLILPLARCKOVFLVGDPKOLPATVISTAOOSGY
GTFMERDOKAGYPKMKIOTYRMHPELLISQHKPIVSLILTPMISYFLQIF
LEFNDRSPKQFEGALEDDSDIEAOTRDMHXYRCFQPCFEPDIHSGKFLHGA
TGSRVNIDVEFEVFLIYHRLVMTPELKSLSOLALISPVYQYKTRKDKRKFMECEA
EKVYDINIVYDGFQYVGSNAITLKSPTLKNRILIESEQGNRLEFKYSKPLNPFSEBNLET
MKLIDEMELPDPELTEDSLFVAPYGGDDDFGDDADDDVANAAGD"

Intron
20374..20766
/gene="AT4g15570"
/number=1
exon
20767..20852
/gene="AT4g15570"
/number=2
Intron
20853..20961
/gene="AT4g15570"
/number=2
exon
20962..21012
/gene="AT4g15570"
/number=3
Intron
21013..21255
/gene="AT4g15570"
/number=3
exon
21256..21367
/gene="AT4g15570"
/number=4
Intron
21368..21466
/gene="AT4g15570"
/number=4
exon
21467..21572
/gene="AT4g15570"
/number=5
Intron
21573..21652
/gene="AT4g15570"
/number=5
exon
21653..21750
/gene="AT4g15570"
/number=6
Intron
21751..21988
/gene="AT4g15570"
/number=6
exon
21989..22117
/gene="AT4g15570"
/number=7
Intron
22118..22181
/gene="AT4g15570"
/number=7
exon
22182..22274
/gene="AT4g15570"
/number=8
Intron
22275..22624
/gene="AT4g15570"
/number=8
exon
22625..22678
/gene="AT4g15570"
/number=9
Intron
22679..22956
/gene="AT4g15570"
/number=9
exon
22957..23015
/gene="AT4g15570"
/number=10
Intron
23016..23121
/gene="AT4g15570"
/number=10
exon
23122..23206
/gene="AT4g15570"
/number=11

Query Match 9.7%; Score 199.4; DB 8; Length 195068;
Best Local Similarity 52.6%; Pred. No. 5.3e-30;

Matches 774; Conservative 0; Mismatches 556; Indels 141; Gaps 10;
QY 7 ccaaaaaataaaaaaaatgacagagctcagctcctctgtgccaattccagacaca 66
DB 4995 CAACAACAAATTCMACTCTCCACCGGCTCCACACTTATTCGTAATCTCCAGCCCA 4936
QY 67 gggacacataatcccgcccttaattgcgccaagcttcg-----caaatgcagcat 120
DB 4935 AGGTCAATCAACCACTCTCGAGTACGCAACGCGCTGCGGCAACAATCTCGGTGC 4876
QY 121 tcaagtcacattcttcacagcgctcagcatgagcgacatgltccaaacgcgcgtg 180
DB 4875 TCGAGTCACTCCCGCCCTCAATCTCTGCGTCAACACCGCCGATGTTCTGACAGAAA 4816
QY 181 ctcaaaagcgctcatcaattgtgtgttttcgacgggtgtgacgggttaagcc 240
DB 4815 CGTCCCGAACCCTAATCTTCCTTCCCTACCTACCTGAGCGCCAGCGGTTCCAAATC 4756
QY 241 cgg-----agaagatggaagaaactacatgltcgagatgaa 276
DB 4755 CTCTGCTTACTCCGCAAAATCTGCTCAAGACGCCACTGGAACCTTCATGTCGAGATGAG 4696
QY 277 aagcagaggtataaagccttgagcgatactcttgacccaataatgtcgataaaaa 336
DB 4695 ACGAGGTGCGAAGAGACACTAACCGAATCAATCGAAG-----ATTAACGGGAACAAA 4642
QY 337 cagcaaaatcacgctcggtggtgacccacccctcttgatgagcgcaaggtgagcg 396
DB 4641 CAGGCTTTTACTTTCGCTGCTTACAGAGATTCCTCACTTGGTGGTGAAGCTAGCGCG 4582
QY 397 tgaattccatctccggagcgcgctactctggaatgagccagctacgggtgttgatata 456
DB 4581 TGAGTTTCAATCT 4522
QY 457 ttaacttatttcaacgagctatagcagcaaaatcgatgaggtgttcgagatgc----- 507
DB 4521 TTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4462
QY 508 taticatctgcggagagatcccaatgctggtgacagcttgatggttaccgttttctct 567
DB 4461 TTCTATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4402
QY 568 tccctccacgcatgagagatccgttcaac---tgatgaaggaagaaatggaacttga 624
DB 4401 CTCTTCMAATGTCTACCGGCTTCTTCTACCGCGGTTTCAGAAAGATTGACTCTGAA 4342
QY 625 aggtgaagaaaaaacctaagtccttggtgaacagcttggtgttgagccctgtagcgt 684
DB 4341 GGAGAAATTAACCTTAAGATCCCTCATCAACACTTTCNAAGAGCTTGAGCCAGAGCCAT 4282
QY 685 caagggcatlgaataagtaagatagatlgaatcagtcggtccgttgatctctccgaatc 744
DB 4281 GAGCTCGGTTCCAGATTAATTTCAAGATGT-----CCCTGTC 4245
QY 745 ggaaggttaagatccttcggaacaggtccttcggcggaagattgttcggaagaggtcg 804
DB 4244 GGTCCGTTACTTAACGTTGAGAACGGAATTT-----TCGAG 4210
QY 805 tgaagcagattgctcctgaatggttgagcaagaaatccctgactcttcgtgtgttcaagtt 864
DB 4209 TCGCGGTGAATACATAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4150
QY 865 gttcggaaagcttgtaatacgaacgaagtcgcaaatggaagagataagcaagaggtcgt 924
DB 4149 GTTCGGGACGCTTGCGCTGTTGAGCAAGAAACGCTTGAGGCTTTGTAAAGCGTTGAT 4090
QY 925 agattgtcgagagcgcttcttggtggtggtgagag----- 959
DB 4089 ACAAGTGGGAGACATTTCTGTGGGTGATTAAGGATTAAGTGCACAAATTAAGAAAGA 4030
QY 960 -----taagcgaaggaagaaaggtatgataagttcgaatgagaggttgaacgaagtg 1014
DB 4029 TGAGCAAGAGAAAGAAAGAAAGATTGATTAAGTGAAGTTCAGAGAAAGCTTGATGAGATAG 3970

QY 1015 gaaaatctatcttggtgtcttcataatggaagtcctgacgcacccctctgtggatgttt 1074
 Db 3969 AATGCTGTTTCATGCTGTCATCAGTTTGGGTTTGAATCATATCATGATAGGTTGTTT 3910
 QY 1075 cgtacacacactcggcggtggaattcgcacttagaagatatacttccggggttcgcgaatggt 1134
 Db 3909 CGTACACCATTTGGCGGTGGAACCTACGCTGGAGAGACTGTTTTCAGSAGAGTCCGGTGT 3850
 QY 1135 ggccttcccgcaatggtgtctgatacgaaggcaatcgcgaactcgtatggaagatctgaggag 1194
 Db 3849 GGCCTTTCGCCAATGAGAAATGATCAGATGATGAACGCGAAGCTTTTGAAGATGTTGTGAA 3790
 QY 1195 gaacggatcgagagtcgagagctaataagga-----gggtagcgtcgtctgatagttga 1245
 Db 3789 AACAGGCTGTAAGACTGATGAGAAAGAAAGAAAGAGAGAGTGTGGTGGATAGGA 3730
 QY 1246 tgaactaagagatgatatctgagagagctatgatatgagtgaggaagaagtaggaacttag 1305
 Db 3729 GGAGATACGGCGGCGGATTTGACGAGAGCTTATGCA-----AGACAAGCGCGAGAGGTTTAC 3676
 QY 1306 agagagtcgctgcgaatggaagattcggcaagaagaagctatcgagagagaatgcatcttc 1365
 Db 3675 AGGAAATGCCACGAGGTGGAAGGATTTACGCGCGAGGCTGTGAGAGAAAGAGGCTCTTC 3616
 QY 1366 agtatacaactcaaggctctctctgatatgag 1396
 Db 3615 CTTTATCATCTCAAAAGCTTTGTGATGAG 3585

RESULT 9 ATCHRIV41/c

LOCUS ATCHRIV41 197419 bp DNA PLN 16-MAR-2000
 DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment NO. 41.
 ACCESSION AL161541
 VERSION AL161541.2 GI:7268270
 KEYWORDS
 SOURCE

ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis

REFERENCE
 1 (bases 1 to 197419)
 EU Arabidopsis sequencing project.
 AUTHORS
 TITLE Direct Submission
 JOURNAL

COMMENT
 Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lmu@chem.mpg.de, mayer@chem.mpg.de, project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@jic.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/Proj/thal/>
 this fragment has an overlap with ATCHRIV40 at the 5' end and an
 overlap with ATCHRIV42 at the 3' end.

FEATURES
 source
 location/Qualifiers

1..197419
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 21638..23498
 /gene="AT4g15240"
 complement(21638..22378)
 /number=1
 complement(join(21638..22378,22677..23498))
 /note="similarity to hypothetical 66.0 Kd protein
 F47D12.7; contains EST gb:AA712915"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CA78566.1"

/db_xref="GI:7268271"
 /translation="MNQTRKRHRPDSLSRLDLSLSPENLFPANSPSLTSLRINIFLL
 LIFCFITIIIFSYGTNREEDLSISRLSVFSTRRRHLLESLAASHDSYLRRSYVR
 LMTSPSTAAVFLDRGGLIEDLPLPVIYVKDVSRRPYNPGRLSALIKAVRIVET
 VDRGDKVMMFVDDDDIVFEDNLVYLSKDYBRRKMFVYGSNEEFIDQAVRSFMA
 FGGGFAISLAKVLAKVILDSCLMKRYSHMGSPSRIFSCVAELGYVLTREPFQID
 VRGNGILCAHPLDSPILSHLDLVDPFPFKRRRTESVAHLIGAASFDGRILQDSV
 CYSLMTYVYVYMGVAVQVYEGKYLPLDLTLOKTSYRGRSGQSNVMSPTREXP
 RDPGRPLVYFPLDSVSGDETBECTNSNTIIHRYGCHRAEAVERLERIVLSPLENRY
 EYGFINTRFSVKVPSTILPGFTFFFGFLKIKQMLPRROCCDISSPVNKSVMINR
 QCPDELILAMNT"
 complement(22379..22676)
 /number=1
 complement(22677..23498)
 /number=2
 28265..29226
 /gene="AT4g15250"
 complement(join(28265..28651,28738..29226))
 /note="similarity to CONSTANS protein, Arabidopsis
 thaliana"
 /codon_start=1
 /product="CONSTANS like protein"
 /protein_id="CA78567.1"
 /db_xref="GI:7268272"
 /translation="NPARCDPGCTFERALTYCKSDSAKYLINCDYNVHNSAPLSORHTR
 SLCEKCSLOPTAVHCHMENVSICGCGQVATSNCTGLRHQLQSLNPTSDCPSPSDGK
 IMESTLEPSYTVLVSFSDTLQEDLDNMGSDVSIVTOTLKLDTSSFFPESNLPRVY
 IEESGGLDCEGINLDDAPLNFNASNDITGSSLDNTKCYEDDSFKENNGILPSL
 LPLPLGNVVPNNLSLMSNLTGESNATDYODCGISPGFLIGDSPWESVNSFNPFLR
 DEAKRYKOKSKRM"
 complement(28265..28651)
 /number=1
 complement(28652..28737)
 /number=1
 complement(28738..29226)
 /number=2
 31839..31950
 /gene="AT4g15260"
 /number=1
 31839..33264
 /gene="AT4g15260"
 join(31839..31950,32018..33264)
 /gene="AT4g15260"
 /note="strong similarity to UTP-glucose
 glucosyltransferase; Contains UDP-glucosyltransferases
 signature MA319-362; contains EST gb:T20820, T46737,
 A1993247.1, A1994260.1, A1996826.1, A1997955.1, T76043,
 AA720097, A1997288.1, W43072"
 /codon_start=1
 /product="glucosyltransferase"
 /protein_id="CA78568.1"
 /db_xref="GI:7268273"
 /translation="MKRIELVTFIPSGIHLNPTVYLAQKLLIGSEEHFNHOGDRLR
 YEPISVEKOPPTANLEPSQVYIEKQPOVRYVSHIINPTKRLGPFVDDFCSSMDIR
 ANEGVQCVITVYSNATFLGIHLHOEMVDKXVSDLSDELVEFLPCLTPRPVYK
 LHIPLISSKMLPEFPAOGRSFRKKMGIIVNTVALLEBHALKMNEDVLAQAVPGV
 LILNDIGDDDEKLELYLRLDDOPKSVYLCFGSGGFEDEORREAVVALVNSGHFE
 LMSLRASPRIMKRPQDYKNLEVLPGDLERTIDRKKVYGADPOVAVLEXPALIGE
 VTGCRSHLESIMFGVPMYTPVLAEOKVANFVETELGAYIRKICISGDLILIGE
 MEIYAEHIERALRCVMEQSDSVRSVKEAKCHVALMDGSSKTLQKFTIDVLEN
 VA"
 31951..32017
 /gene="AT4g15260"
 /number=1
 32018..33264
 /gene="AT4g15260"
 /number=2
 34599..34620
 repeat_region
 complement(34923..35221)
 /number=1
 34923..35916
 /gene="AT4g15270"
 complement(join(34923..35221,35280..35916))

/note="similarity to UTP-glucose glucosyltransferase;
contains EST gb:T20820"
/codon_start=1
/product="UTP-glucose glucosyltransferase like protein"
/protein_id="CAB78569.1"
/db_xref="GI:7268274"
/translation="MKIQLVFIPSPGIGHRSTVEAKRIYNSDRRLMTVITIPPS
ISDDVEITYIASLTASQDLNLEATSVANQPEPDYOEPTVTEKQKQVDRVARI
FHSSTGDSPRVAGFVMDPCSMIDVNEFGVCYVNTSNACIGITLHTIDRMDK
KYVSELDSVNELEFPFLTRPVKCLPDFEFTKMLAEFLAQCARKMAYVWPL
YAEOKISAFVAVEELGAVOIRKFRGDMVLCGMEIVTTVDIERAVRCVENDSEVRN
RVKEMAEKCHVASMDGSSQVYALOKFIDVTENV"
/complement(35222. .35279)
/number=1
/complement(35280. .35916)
/number=2
37302. .38738
/gene="AT4G15280"
/number=1
37302. .38738
/gene="AT4G15280"
37302. .38738
/note="strong similarity to UTP-glucose
glucosyltransferase; Contains UDP-glucosyltransferases
signature AA346-389; contains EST gb:T20820, T46737,
A1993247.1, A1996826.1, T76043, AA720097, W43072"
/codon_start=1
/product="UTP-glucose glucosyltransferase"
/protein_id="CAB78570.1"
/db_xref="GI:7268275"
/translation="MKIELVFIPPLPGIGHRPVYAKLOIGSENFSLTITITISRP
DAGDSACIASLTASQDLNLEATSVANQPEPDYPAVYIEKTKTVRQVAV
ANIVPTKRLAGFVMDPCSMIDVNEFGVCYVNTSNACIGITLHTIDRMDK
YVSELDSVNELEFPFLTRPVKCLPDFEFTKMLAEFLAQCARKMAYVWPL
ALDEPHALKMENINGDDLPQYVPGVPLHLEKNDDEKSELRLMDQPSKVPL
CGSLGFTPEBOTRETNALDRSGRFLMCLRAHSPNKRDRDRLTLEVLPGLF
ERTLDRGVIMAPQAVLEKPAIGFVTGCSNLTSELMCEVPATVPLAEQKVN
AFEMTEGLAVETIRKYLKGDTPAGMEYTAADIERAIRVMEQSDVIRNNKEMAE
KCHFLMDGSSKALEKFDVTEEND"
/complement(39813. .40234)
/number=1
39813. .43103
/gene="AT4G15290"
/complement(join(39813. .40254,40297. .40391,40490. .40837,
40928. .41106,41541. .41676,41753. .41973,42204. .42366,
42409. .42663,42810. .43103))
/note="similarity to cellulose synthase, Agrobacterium
tumefaciens; Contains Prokaryotic membrane lipoprotein
lipid attachment site AA320-330; Prokaryotic membrane
lipoprotein lipid attachment site AA554-564; contains EST
gb:A1999203.1"
/codon_start=1
/product="cellulose synthase like protein"
/protein_id="CAB78571.1"
/db_xref="GI:7268276"
/translation="MVFPTADTVRESPIITVNTVLSLLAVNPANKIACVSDGCS
LTVPSLKASKRVKIMAPFCKYVNRAPRRFLNPLVATDSVFSKMKMKLYKV
FYVYFECIMNRKRYKEDATGSDHLDADDPEAFSNTKPMDSHTIVKILKL
FKTAYRVQFQSKVMYILKLIIVMENKGVGDEKEVPLVYISREKRPNTLHYKT
GAMNPLVNRKLSHTEFEFLRLMTNAPTLVNDPMVANEEDVAVRCPFLNLSQNR
HCAVQFQKRYDYSTNELAVIOSVFLFLDPEFLIGRGVATGCGCHRR
RMYGLSDSDLEDNKNISQVATREFLADESLYKRGSKRLVASYDALQKRSNPKS
LANLEAAGEVGHCHETQISWGNLGMHTVSADRLITSGILKRWISFISPDPA
FGSTPTGLDEIVQQRWATGALIEVLNQSPPMGFPHGRKIFKORLRYEVALDLR
S1PELIYCLIPAYCLHDSALFPGPKICITVYLVGNAHSLMDQPSKVPL
FSIODIILKILGISOIGFVIAKKTIPETKSVYESKPSQGEDVPRKNTLKRFSDSG
LPIFGFTMLVNLAAAGLYVLRLOSCSHGGSGSGLAEACGCLVYMLPLPLKGLFE
HKYSIPLSTLSKAAFLVLFVFCVCK"
/complement(40255. .40296)
/number=1
/complement(40297. .40391)
/number=2
/number=2

Query Match 9.7%; Score 199.4; DB: 8; Length 197419;
Best Local Similarity 52.6%; Pred. No. 5,3e-30;
Matches 774; Conservative 0; Mismatches 556; Indels 141; Gaps 10;
Intron
complement(40392. .40489)
/number=2
exon complement(40490. .40837)
/number=3
intron complement(40838. .40927)
/number=3
exon complement(40928. .41106)
/number=4
intron complement(41107. .41540)
/number=4
exon complement(41541. .41676)
/number=5
intron complement(41677. .41752)
Query 7 ccaaaaaataaaaaaataggacagagctcgtcctctgagccatcccaagaca 66
Db 197414 CAACACAAATTCACACTCCACCGGTCACACTTCTATTCGAACATTCACGCCA 197355
Qy 67 ggacacataatcccgccctcaatctgcaagctctg-----caatgcgaat 120
Db 197354 AGGCACATCAACCCATCTCTCGAGTCACCAAGCCCTCCCGAACAAATCTGTGTC 197295
Qy 121 tcaagtcacatcttcacacagcgtctacgacatgagcgagcgtatgccaagcgcgtcg 180
Db 197294 TCGAGTCACCTTGCGCGCCGTCACATCTCTGCGTACAAACCGCGCATGTTCTCTACAGAAA 197235
Qy 181 ctcaacagggctcatcaattgctgctcttcgcagcggtatgaagcgggttaagcgc 240
Db 197234 GTCGCCGGAACCCATATCTGCTGCTACCTACCTCCATGCGCACGACGAGTTCAATC 197175
Qy 241 cgg-----agacgatggagaagactacatgctcgagatgaa 276
Db 197174 CTCGCTTACTCCGACAAATCTCTGTCACAGCCGACGGAACCATCATGCTGAGATGAG 197115
Qy 277 aagcagaggtatataaagccttgagcgaatcctctgcagccaataatgctcaaaaa 336
Db 197114 ACGACGTGGGAAAGACACTAACCGAATCAATCAAG-----ATAACGGAAACAAA 197061
Qy 337 caacaaatcacgtctgctgctgagccatcttgcattgagcgagcgcaagctgagcgc 396
Db 197060 CAGGCTTTTACTGCTGGTGGTTACACGATCTCTCTCACTTGGGTCGTGAGCTAGGCGC 197001
Qy 397 tgaattccatctccgagagcgctactctgattgtagcagcactacggtgtgtgatatc 456
Db 197000 TGAATTTACT 196941
Qy 457 ttactttattcaacgagctatagcagcgaatcgatcgaggttcgagtc----- 507
Db 196940 TTACCAATTAATGCGTACGACGATGCAATCTCAGAGATGGTATACCCCTCTAG 196881
Qy 508 tattcaattgcccgagagactccagctgctgagccagcggtatctacgcgtcttccctc 567
Db 196880 TTCTATTAAATTAACCTTCTCTGCACTGCTCTCTCTCTCTCTCTCTCTCTCTCT 196821
Qy 568 tccctccacgcataagagaltccgttcaac-----tgaatgaaggaatlgaaaattaga 624
Db 196820 CTCCTCCAAAGTCTACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 196761
Qy 625 aggtgaagaaaaaacctaaagctctgctgtagaagccttgcgtctgtagagcctgagc 684
Db 196760 GGAAGAAATTAACCCCTTAAGTCTCTCACTCAACCTTCTCAAGAGCTGACGAGGCAT 196701
Qy 685 caagcactgataagtagacagatgtagtcaatcgagcgttgcgttgccttcgcatctc 744
Db 196700 GAGCTCGTTTCCAGATTAATTTCAAGATGT-----CCCTGTC 196664
Qy 745 ggacggtataaagaltccctcgagcaggtccttgcggagagattgttcgagaagagtgaa 804
Db 196663 GGTCCGTTACTCAAGTTGAGAACGGAATTT-----TCGAG 196629

intron
9800..9224
/gene-"d13768w"
/number=2
9225..9395
/gene-"d13768w"
/number=3
9396..9520
/gene-"d13768w"
/number=3
9521..9574
/gene-"d13768w"
/number=4
9575..9653
/gene-"d13768w"
/number=4
9654..9807
/gene-"d13768w"
/number=5
complement(10065..10511)
/note="similarity to L.esculentum Tfms gene"
/codon_start=1
/product="glycine-rich protein like"
/protein_id="CAB10324.1"
/db_xref="GI:2244903"
/db_xref="SPPRMBL:023398"
/translation="KQITRTAHRYVTNMAKLAVLTITLLVNCITYYAKGGGGAH
GGGVGHVSGGAHSVGGHAGGGGHAASCGGCHAAVEGGGHGGGGGHE
EEGGHGIGRGGAHVHPATRNNGSTLTIPVDGSTVFMTIFFLN"
complement(10065..10511)
/number=1
10065..10511
/gene-"d13770c"
join(12306..12597,13233..13495,13554..13835,13944..14000,
15041..15290,15396..15602,16095..16280,16380..16453,
17015..17346)
/gene-"d13775w"
/note="similar to hypothetical protein C02P5.7,
Caenorhabditis elegans; contabs EST
gb:12137311;937779;315469"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10325.1"
/db_xref="GI:2244904"
/db_xref="SPPRMBL:023399"
/translation="MDKPFGTASVSMGSDRSACKDIDEMVGENTLYPGISTENO
LRKGFRKAYGLISADLIITLISAVVLANPVNDLTGSPGILPLVCIVPELTSFE
FVSTEFLDVLSERTSAIEMSYLETRPIAHVIDWEFNVPVIMPLATHIOKHNPVNL
LMLFVSLFSFYWGSCAMTEGCYLPLINSFYCHKEENFTWNSKACSKCVTNML
FLRLRYLPGLVIQALITLLSYVGSLETPYPMARKKDPSFGPLTFSLIIIVYS
FIOMPEPLGTVAIVYGFSALKRRRGDDSSPSSRRKLLDTKTHSGAENESSIF
DTGITLANGPRNTLSLIMCPNVSSGICSLAQKTSLKSLDGCCVGEQGIAAV
GRCKOLEPENSIRECGELTDVVIDVYGCSSKLSKTGVAASKATTDELSVSHCK
LLEYLVDSERYIHDKGLIAVAGCGRKLNNLKQCYSTDYAFAGELDTSERIALY
SPHFIDGSRKAIGESKKLADLTDCYFPVSCGLEAI AHGCKELENERINGCNIG
TRIEIAGSCRELALALYCQRIGNSALOIEIKGLAKGPTDRFONIGMPDLIEL
GECPCPKDIIVSHCHITDNGLNIHLVOCKCLLETCHMYCPGITSAGAIVSSCPH
IKKVLIEMKKEVERTERRAGSVISYLCMDL"
12306..12597
/gene-"d13775w"
/number=1
12306..17346
/gene-"d13775w"
12598..13232
/gene-"d13775w"
/number=1
13233..13495
/gene-"d13775w"
/number=2
13496..13553
/gene-"d13775w"
/number=2
13554..13835
/gene-"d13775w"

	Intron	/number=3 13836..13943 /gene=*d13775w*
	exon	/number=3 13944..14000 /gene=*d13775w*
	Intron	/number=4 14001..15044 /gene=*d13775w*
	Source	/number=4 14700..47605 /organism='Arabidopsis thaliana' /db_xref='taxon:3702' /clone='cosmid YUP5B11.IBIIL2/YUP5B11.IBIAB'
	exon	/number=5 15045..15290 /gene=*d13775w*
	Intron	/number=5 15291..15395 /gene=*d13775w*
	exon	/number=5 15396..15802 /gene=*d13775w*
	Intron	/number=6 15803..16094 /gene=*d13775w*
	exon	/number=6 16095..16280 /gene=*d13775w*
		/number=7
Query Match		9.7%; Score 19.4; DB 7;
Best Local Similarity		52.6%; Pred. No.5,3e-30;
Matches 774; Conservative 0; Mismatches 556; Indels 141; Gaps 10.		
QY 7	ccaaaaaataaaaaaaaaatgagcagaagtctacgtcccttggcacattcccagaca	66
DG 47941	CACACACAATTCCAACTCTCCACCAGTCCACACTTTGTATTGCATAAATTTCCAGCCA	47882
QY 67	ggagacataaataccgcacctcaacttcgccgaaggctcg-----caaatgcagcat	120
DG 47881	AGGTCAATCACGCCATCTCTCGAGTGACGAACGGCCTCGGGAGAACAATCTGTGTGC	47822
QY 121	tcaagtacattctttaccaagsgtgtcagaacgagcgcatgltyccaagaacgcgttg	180
DG 47821	TGAAATCACTTCCCCTCCCATTAATCTCTCCTACAACCGCCGCATGTTCTTACAGAAA	47762
QY 181	ctcaaacgggtcatacaattlbtlygcgtlttcgacgggatbaigacagggittacagcc	240
DG 47761	CGTCCCCGAAMCCCTAATTTCTGCTACTACTCGATGCGCACGACGAGGTTTTCAATC	47702
QY 241	Cgg-----aaagatatgggaagtaacataatgttcgagatgaa	276
DG 47701	CTGTGCTTACTCCGANAAATCTGCTCAAGAGCCCACTGGAAAACTTCATGTCTGAGTAG	47642
QY 277	aagcagagtataaaagcctitgtagcgtatcctitlgcaagcaaalatigtalcaaaaaa	336
DG 47641	ACGACGTGCGAAAGAGACACTAACCGAATTAATTCGAG-----ATAACCGAANAANNA	47588
QY 337	cagcaaatcaagtcgtgtgtgagtlgtacccccaccttcttgatgggcgagcaagtgagg	396
DG 47587	CAGGCTTTTACTTGTGCTGTGATTACAGATTTCTCCTCACTTGGGTCGAGCTAGAGCGC	47528
QY 397	tgaattcalatlcgcgagcgcgclactlctgatltagacgaactacgytltgtatalat	456
DG 47527	TGAGTTTCACTTCCTCTCTCTCTCTTTGGGTCCCAACAGTAACAGCTTCTCATTTW	47468
QY 457	ttaattlatttaacgcgctatagcagaabaatgalcygtgcgagalg-----	507
DG 47467	TTACATTACTTCAITBGCTACGAAGATTCATTCADAGATGGCTAATFACCCCTTAG	47408
QY 508	tattaactgcggagagactccagtgcttgccagagcgtgattaccgcttccctct	567
DG 47407	TTTCATTAATTAATTAATCTTCCTGCGACAGTGTCTGTGCTGCTGATATATTCCTTCATTTG	47348

```
OY 568 tccctccacgcacatgagagatccgcttcac---tgaagaaggagaaattggaacattaga 624
      |||||
Db 47347 CTCCTCCAAATGCTACGCGCTTCTCTACCGCTTTCGAGAACACATTGATTCACATGAA 47288
OY 625 aggtgagaagaanaaaactaaaggtcttgtagaagacttgtagcgttgtagcctatgcgt 684
      |||||
Db 47287 GGAAGAAATAAACCCTTAGATCCATCAACACACTTCCAGACACTTGCAGCAAGACCAAT 47228
OY 685 caaagcaatctgaatgaatgcagaagatctgcaactcgagcgcttgatccctccgactcct 744
      |||||
Db 47227 GACCTCGGTTCCAGATTAATTCACAGATTGT-----CCCTGTGC 47191
OY 745 gtagcgttaagaatccctcgcgaaggtcttcgcgcgaagatctgtcgaagaaggcgtcga 804
      |||||
Db 47190 GGTCCGTTACTAACGTTGAGAACGATTTT-----TCGAG 47156
OY 805 tgaagcagatctgctcgaatgtgtgtagcagaacatccctcgtctcgttgtagcttc 864
      |||||
Db 47155 TCGCGGTGATACATAGAGGTGGTGGATACATAAGCGGATTCGTCTGTATGTTTC 47096
OY 865 gtccgaagactcgttaatacgaagaagtcgaatggaagatagaagaagagcgtct 924
      |||||
Db 47095 GTTCGGGAGCGTTGCCGCTGTTGAGCAACAACAGCTTGTGAGCTTGTAAAGCGTTGAT 47036
OY 925 agatctgtagaagccgcttcttgtagtgtagaag----- 959
Db 47035 ACAAACTCGAGACCATCTCTGTGGTGATTACGGATTAAGTCGTACAGAAATAAGACAGA 46976
OY 960 -----taaaacgaaggagaagagatctgataagctgcagagagctggaacgaagtcgg 1014
      |||||
Db 46975 TGAGCAAGAGAGGAAGAAAGATTGTCATTAAGTTTCAGAGAAAGCTCATGAGATAG 46916
OY 1015 gaaatctgatacttggtcttcctcaatggaagctcctgaacccctcgttgtagctt 1074
      |||||
Db 46915 AATGCTGCTTATAGTGTGATGATTAAGGCTTTTGAATCATATGATCGATAGGTTGTT 46856
OY 1075 cgtgacacacatcgagtgtagatctgactcctagagagatatactcctcgagtgtagt 1134
      |||||
Db 46855 CGTGAGCGCATTCGCGGTGAGACTACCTGAGACGCTTGTTCAGAGATTCGCGTGGT 46796
OY 1135 ggccttcgcagtggttcgactcaagsgagacgagctgagatgtagagatgtgtag 1194
      |||||
Db 46795 GCGCTTCCGCAATGGAATGATGATGATACACCAAGCTTTTGAAGATTTGTTGAA 46736
OY 1195 gacgggtctgtagagtagagctaatgagga-----ggtagcgtctgtagtga 1245
      |||||
Db 46735 AACAGGTGAAGAGTATGAGAGAAAGAAAGAGATTGTGTGTGATAGTGA 46676
OY 1246 tgaatctgagagatgtagagaggtttagatgtagtgtagagagagtagagactgag 1305
      |||||
Db 46675 GGGGATACGCGCGGTGATGAGAGATTATGA-----AGACAAAGCGGAGAGAGTTTGA 46622
OY 1306 agagagatctgtagcagtgtagagatcttgcaagaagaagctatgtagagagatgtagt 1365
      |||||
Db 46621 AGGAATTCGACAGAGGTGAGAGATTAGCGCGGAGGCTGTGAGAAAGACGCTTTC 46562
OY 1366 agttaaacaaccccaaggtcttctgtagag 1396
      |||||
Db 46561 CTTTAATCATCTCAAGCTTTTGTGATGAG 46531
```

```
RESULT 11
ATU81293 1589 bp mRNA PLN 02-JUN-1997
LOCUS Arabidopsis thaliana UDP-glucose:indole-3-acetate
DEFINITION beta-D-glucosyltransferase (laglu) mRNA, complete cds.
ACCESSION U81293.1 GI:2149126
VERSION 1
KEYWORDS thale cress,
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (bases 1 to 1589)
AUTHORS Graham,R.A. and Thornburg,R.W.
TITLE DNA Sequence of UDP Glucose:Indole-3-acetate
BETA-D-Glucosyltransferase from Arabidopsis thaliana (Accession No.
U81293) (RGR97-044)
JOURNAL Plant Physiol. 113, 1004 (1997)
2 (bases 1 to 1589)
AUTHORS Thornburg,R.W. and Graham,R.A.
TITLE Direct Submission
JOURNAL Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State
University, Ames, IA 50011, USA
FEATURES
source
1..1589
location/qualifiers
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
26..1589
/gene="laglu"
26..1450
/gene="laglu"
/function="transfers the glucose moiety from UDP-glucose
to indole-3-acetic acid to form ester linked IAA-Glc"
/note="IAA-glucose synthase"
/codon_start=1
/product="UDP-glucose:indole-3-acetate
beta-D-glucosyltransferase"
/protein_id="AA58497.1"
/db_xref="GI:2149127"
translation="MANNNSPTGPHPLVTPFPAQHINPSTELKRLAGTISGARY
TPASISAYNRMRSTENVPETLIPFTYSDGDDKSSAYSDKROATNPFSEMR
RKGETLLEIEDNRKONRPFYVVTITLTWALERERFLPSALWQVYTPSTF
HYFNGEDALISEMANTPSSILPSPLIVADPSFLVSSNVAFILPAFRQIDS
LKEINPILINTFOLEPEPAMSSVPDNKRIVVGLTLTRDFSSREYLEMDTGA
DSVLVYFGLAVLSRKQLELKAQISRPFILWITDKSYRNKDEQREDCIS
SSEKSFDEIGMNVSMCDQFVILNHRISIGFVTCGNSTLESILVGVVAFPMQNDQ
MTNKLLEDCKTQTVRVEKKEEGVYVDSSEIRRCIEVEMDAEERFGRNATRWKD
LAAEVRREGSSFNHLKAFVDEHM"
polyA_signal
1568..1575
/gene="laglu"
polyA_site
1588..1589
/gene="laglu"
/note="24 A nucleotides"
BASE COUNT 433 a 342 c 375 g 439 t
ORIGIN
```

```
Query Match 9.6%; Score 197.8; DB 8; Length 1589;
Best Local Similarity 52.5%; Pred. No. 1.1e-29;
Matches 773; Conservative 0; Mismatches 557; Indels 141; Gaps 10;
```

```
OY 7 ccaaaaaataaaaaaatgagagagcagctcctctgagcaccatccacacaca 66
      |||||
Db 31 CACAAACAATCCAACTCCGACGGTCCACACTTCTATGTAATTTCCACGCCA 90
OY 67 gggagacacaaatccgcctcctcaatctgcgaagcgtctcg-----caatgccaacat 120
      |||||
Db 91 AGGTCAATCAACCCATCTCTGAGCTAGCCAAAGCCTCCCGGAACAATCTCTGTGC 150
OY 121 tcaagtcacatcttcacagcgctcagcagcagcagcagcagcagcagcagcagc 180
      |||||
Db 151 TCGAGTACACTTCGCGCGCTCAATCTGCTACAAACCGCGCATGTCTTCACAGAAA 210
OY 181 ctaaacagggtccatcaatcttgctgcttcgcgaaggtatgaagagaggttaagcc 240
      |||||
Db 211 GGTCCCGGAACCTTAATCTTGCTACTACTCGGATGCGCACGACGCGTTTCAAAATC 270
OY 241 cgg-----agacgatgggaagaaactacatctgcgagatga 276
      |||||
Db 271 CTCGTGCTTACTCGAACAAATCTGTCGAAGACGCCACGAAACCTTCAATGTCGAGATGAG 330
```


OY	277	aagcagagggataaaagcccttggagcctactctctgcagacgaataatgtcgttcaaaaag	336
Db	331	ACGACGCTGGCAAGAGACACTTAACCGCACTAATCGAG-----ATAACGGAAACAAA	384
OY	337	cagcnaaaatcaagcttcggtgtactcccaactctttgcatagtgcgcgcaaggtgagcg	396
Db	385	CAGCGCTTTTACTGCTGGCTGTTTACACGATATTCCTCCACTTGGTGGCTGAGCTAGCCG	444
OY	397	tgaattccatctccggagcgcgctactctcgatattgagcagactagagcgttgatalt	456
Db	445	TGACTTTCATCTCTCTTCCTCTTCTTTGGGTCACACGATACAGCTTCTTCATTTT	504
OY	457	ttacttlatlctcaacgagctatagcagcaaatcgatgcggttctgagtc-----	507
Db	505	TTTACCATTAATCTCAATGAGCTACGAAAGATCTCAGAGAGGCTTAATACCCCTTAG	564
OY	508	tattcaactgcccggagagatcccaagtgcctgcgcccagtgatttaccgcttccctt	567
Db	565	TTCTATTAAATTAACTCTCTCTGGCACTGCTACTGCTGATTCGGTATTTCTTCTTCATTTG	624
OY	568	tcttccacagcatgagagatccggtcac---tgatgaagagaaattggaacttga	624
Db	625	CTCTCTCAATCTCTACCCGCTTTCTTTACCCCGGTTTCAGAAACGATGTGATTCCTGAA	684
OY	625	aggtgaagaaaaaacctaaagttcttggtgaacagcttgatgtgcttggagccatgagct	684
Db	685	GGAGGAATAATTAACCTTAAGATCTCATCAACACTTTCCAAAGAGCTTGAGCCAGAAAGCAT	744
OY	685	caaggtcatttgataaagtagagatlttgatgcaatcgggcgcttgatcttcccgaaactt	744
Db	745	GAGCTCGGTTCCACATTAATTTCAAGATGT-----CCCTCTC	781
OY	745	ggacgtgaagaatcccttcgcagcagcgtcttcgcgagatltgttcgagaagagctgaa	804
Db	782	GGTCGCTTACTAAGCTTGGAGAACGGATTT-----TCAG	816
OY	805	tgaagcagatattgctcgaataggtttagcacgaatctcgatcttcgtgtgttaccgttc	864
Db	817	TCGCGGTGAATACCTAAGAGGTGGATACCTAAAGCGGATTCGTCTGTTAATGTTTC	876
OY	865	gttcggaagcttcgttaatacgcacgaatctgcnaatcgtgaagaagatagcaagagcgctt	924
Db	877	GTTCGGGACGCTTCCCTGTGTGAGCAAGAACGCTTGTTGAGGCTTTGTAAAGCGTTGAT	936
OY	925	agattgtggagagccgcttlttgtgggtgtgaag-----	959
Db	937	ACAAAGCTCGGAGACCATTTCTTGGGGTGATTACGGTAAAGTGTACAGAAATTAAGAGA	996
OY	960	-----taaacgaagaaagaagttattgataaggtcatgagagaggttgaaacgagtggg	1014
Db	997	TGACAGAGAGAGAAAGAAACAATTTGCATTAAGTACTTAGAAGAGACCTTCGATGAGATAG	1056
OY	1015	gaaaatgtatcttgtgtgtctcaattggaagtcctgcagcaatccctcgttggagatgtt	1074
Db	1057	AATGCTGTTTCAATGAGGTGATCAATTTAAGGTTTGAATCATAGATGATGATGGTTGTTT	1116
OY	1075	cgtgacacactgcggggtggaatccgcatccagagagtatacttccggggttccgattgt	1134
Db	1117	CGTAGCGATGTCGGGTGGAACTCTACGCTGGAGACCTTGTTTCAGAGAGTTCCGGTGTGT	1176
OY	1135	ggcttccgcagtggttcgatacaagggagcaatctcgaaactgtatggaagatgtgtgag	1194
Db	1177	GCGCTTTCCGAATAGGAATATCATGATGAGAACCGGAAGCTTTTAAAGATTTGTGGAA	1236
OY	1195	gacgggtgtgagatctgagacttaactagga-----gggtagcgttcgttgaatgctga	1245
Db	1237	AACAGGCTTAAGACTGATGATGGAGAAAGAAAGAAAGAGAGTTGTGGTGGTGGATAGTGA	1296
OY	1246	tgaaatctagagatgtcttcaagagatltatgatatgtggtggggaagaaagatagaaactag	1305
Db	1297	GGAATTAACGGCGGCTTATGAGGAAGTTATAGA-----AGCAACAGCGGAGAGACTTTAG	1350

QY 1306 agggagtcgtgcgaagctgcaagagatttgcgaagaanaacatlgagaaagatgagatctc 1365
 Db 1351 AGGAATGCGACGAGGTGGAGAGATTTCAGGGCGGAGAGCTGTGAGCAAGAGAGCGCTTC 1410
 QY 1366 agttaaacaactcaagctcttctctgtagag 1396
 Db 1411 CTTATATCATCTCAACGCTTTGTCATATAG 1441

RESULT 12	AC007153/c	LOCUS	DEFINITION
AC007153	103223 bp	DNA	17-MAY-1998
Arabidopsis thaliana	chromosome I	BAC F3F20	genomic sequence,
complete sequence.			

VERSION	AC007153.2	GI:4580365
KEYWORDS	HTG.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	

REFERENCE
1 (bases 1 to 103223)
Arabidopsis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Bryophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 103222)
Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Luiros, S., Schwartz, J., Shinn, P., Tortiuni, M., Vysotskaya, V.S.,
Walker, M., Yu, G., Ecker, J., Theologos, A. and Davis, R.W.

TITLE Walker M., Yu G., Ecker, J., Theologos, A. and Davis, R.W.
JOURNAL Direct Submission
Submitted (24-MAR-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

REFERENCE
AUTHORS
3 (bases 1 to 103223)
Pedersoli, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altai, H., Arzajo, R., Huizer, L., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Tortumli, M., Votoksaia, V.,
Walker, M., Yu, G., Ecker, J., Theologidis, A. and Davis, R.W.
TITLE
JOURNAL
Direct Submission
Submitted (08-Apr-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA (bases 1 to 103223)

4 (classes 1 to 103223).
REFERENCE
 Pederspiel, N.O., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altai, H., Arasu, R., Hutzler, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Lueros, S., Schwartz, J., Shinn, P., Tortiuni, M., Vycotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologos, A. and Davis, R.W.
 Direct Submission
 Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT

On Apr 9, 1999 this sequence version replaced g1:4508084. Bases 1-33654 of clone F3F20 overlap with bases 50550-84203 of 'RAMU' clone T25N20, gblAC005106. e-mail for correspondence: arabseqsequence.stanford.edu Genes with similarity to proteins in the databases are described as 'putative', 'like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genomic.stanford.edu/~chris/GENSCAN.html>), Fexa (Y. Solovoyev

[illegible]

LOCUS	DEFINITION	PLN	05-APR-2000
AC006533	99188 bp DNA		
AC006533	Arabidopsis thaliana chromosome II section 179 of 225 of the complete sequence. Sequence from clones T9H9, F20M17.		
AC006533	AE002093		
AC006533.7	GI:6598600		
HTG.			
thale cress.			
Arabidopsis thaliana			
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
1 (bases 1 to 99188)			
Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Beito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldguy, T.V., Buel, C.R., Ketchum, K.A., Lee, J.-U., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carreira, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Merriam, M.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.			
Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana			
Nature 402 (6763), 761-768 (1999)			
20083487			
2 (bases 1 to 99188)			
lin, X.			
Direct Submission			
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA			

COMMENT

On Dec 17 1999 this sequence version replaced gi:14887240. The sequence and annotation of this chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GENAIL (<http://arthur.asu.ornl.gov/pub/graal/>), GeneFinder (Phil Green, University of Washington), GeneScan (Chris Burge, <http://genome.stanford.edu/GENSCAN.html>), and NecPlantGene (<http://www.cds.dtu.dk/services/MetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curation based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/AI consortium for sequencing PAC clones F6P23, F5J6, T1175, and T13L6, the ESSA group for sequencing clone FL3D, and Scott Jackson, Jiating Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Yixin Zhou, Hanif Khatkhat, Michael E. Henney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tlgr.org.

FEATURES

1. .99188

misc_feature

/note="Sequence from clone T9H9"
complement(<1064. .>2005)

gene

```

/note="T9H9.24; predicted by genscan"
complement(1064..2005)
CDS

```

```
/note="hypothetical protein"  
/codon_start=1
```

misc_feature

```

/note="Sequence from clone F20M17"
complement(join(3588..3973,4058..4094,4189..4265))
mRNA

```

gene

complement(3588. .4265)	
-------------------------	--

CDS

```
complement(join(3813. .3973,4058. .4094,4189. .4260))
```

```
/gene="At2g31730"
/note="unknown protein"
/codon_start=1
/protein_id="AAD32299.1"
/db_xref="GI:4887763"
/translation="MOTDASVLLGMOYIOFLQEQVKCSLASNPYYRVEEYSLRS
KGLCLPLETEVAOTNGADIMAWKPTSSHAFLNSSNSPQ"
/complement(join(<7083, 7286, 7407, 7557, 7647, 7813,
7921, 7995, 8070, 8150, 8243, 8367, 8449, 8698, 9242, 9289,
9386, 9469, 9576, 9833, 9973, 10145, 10486, 10535,
10653, >11039))
/gene="At2g31740"
/complement(<7083)
/gene="At2g31740"
/note="F20M17.22; Predicted by genscan"
/complement(join(7083, 7286, 7407, 7557, 7647, 7813,
7921, 7995, 8070, 8150, 8243, 8367, 8449, 8698, 9242, 9289,
9386, 9469, 9576, 9833, 9973, 10145, 10486, 10555,
10653, 11039))
/gene="At2g31740"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD32298.1"
/db_xref="GI:4887762"
/translation="MGKKGNKAASTDDPLOTLEDFTSKKNMDKFFTLRGNDSEW
YARPOLHSLPLLODSSSSDLOTLVPGCGNSRLTEHLVAGCFRDTITNDESKV
VIEDLRRNIRTRPELRNRNDITKQALADSFTVLQDKGLDMLPEPVALPSER
EGKMNKVNIAQRSLKTYNVAKEKNSVLLHETSAFELVLSGRNDGSGNSRAVL
ESENKIRDCNNGSDLLYSHEDLKIGIDLALETGRRIKETLGGGGSNSYRAVL
DAOKTEPYYCGVFLPKRAHEMLFCSEGGQMOVESQARLIMVLDSHSGA
TMEDIVTSPMTGKVVVEDVYVESAPSNLEDLSTSDLAERLVKFTBEGLISELL
VEDGELEOSOKETKNSQSKRGNKONPSPRLMARVSHDYLAASSYHTLIGSEF
LVSSYLKKAESCGTGWKTVIGLGLLPMFLHGLPEPSTEAVEIDPVMVLSVGIKF
GFPQNRVLVHADIKFTIRDTITNSSESSNSGSDTTHNNGGCPKILLID
VDSADSGCTGCPASDTEFFELISVKRALPOHCLFYNTVYTRSOAKDVMVRMKV
FDHLRLOLEEDENNVVFLGLCSSESVISENDIPESAVTLEGLKCRLETKOSTIDA
TKLKCWK"
/repeat_region
10152..10235
/rpl_family="(7NA)n"
/complement(12661..12688)
/rpl_family="(7AAA)n"
/join(13281..13986,15175..15987)
/gene="At2g31750"
13281..15987
/note="At2g31750"
/join(13342..13986,15175..15900)
/gene="At2g31750"
/codon_start=1
/product="putative glucosyltransferase"
/protein_id="AAD32297.1"
/db_xref="GI:4887761"
/translation="MGEKKNVLYFSPICGHINPLIOESKRLSKNNVTEFLTSS
TSHSILRAITGTGATPLSPFVIDDGEEDHPTSDSPDFAQFQNSVRSISELIS
SMPPKFNAYVDSCLPYLDVCRKHPGYAASFPQOSTVAATYIHLRCEFEFEND
VVLPAPPLKNGIDLPYLDNLICRPLLELISQFVYNDIDFPLVNSPELEVEYV
MKNOWPVKNIGPMIPSNYLDKRLAGDDYGNJFNADVNSCLMDLSDSGSVIYVS
EGSLAVLKDOMEYVAGLKORGNHFLKPSYIEDIDCKLILNMSBO
LOYLAKHSIGCFPHNCGNSTLEALSIVALLGNRPATSDQTNKFTEDYWKGYRK
ADONGFVPEBEIYKCVGEVMEDESEKGEIKRNARLMERAREALSDGSDKNIDEF
VARIVR"
/repeat_region
22837..22997
/rpl_family="(CAAAA)n"
/repeat_region
23089..23130
/rpl_family="(POLY A)"
/join(<24523, 25698, 25968, 26061, 26146, 26420)
/gene="At2g31760"
<24523..26420
/note="At2g31760"
/join(24523..25698, 25968, 26061, 26146, 26420)
/gene="At2g31760"
/codon_start=1
```

```
/product="putative ARI-like RING zinc finger protein"
/protein_id="AAD32296.1"
/db_xref="GI:4887760"
/translation="MDYSDDDIMIDNESENNSDGNESYNNAVDTLILSEKSYV
IIRKEELIKLRDRIERSTLIFLSQVAIVLLYHHYCSKLEDEFTDEERKRV
GLIKEPVDVNGTEVDIOCGIFESYTRKELIYVSCGPYCKTQMTYITTKLEDGG
CLKEPEPVCYAVVGDMDIDVTEKRNKDYRFLRSYVEDKKKKKPSGCCGA
VERGESSEGDVATCYSRCNCSGSDAISPDCEVSMIFKKNDESENKNTLANSK
PCRCRPIKRSKGNHNTCSASCHRCWICGKYSDBHACNNVEADHDKRTILO
SEIRTRTHYKSVNENSRSLKMSDLERFQSVQIKQSDNOCKPKIDLPYDAWLO
IIECRNVLKATYAVGYLDNLKRLPLFEYLGGEATGLERLHCAENELKQFIRKSD
PSDFNAPFRMLKTLGTRKTKYFDNLVYALENLADVTXKSEBSADPLETKLYDAVI
SEGCFF"
/repeat_region
27052..27180
/rpl_family="(TAAA)n"
/join(<27609, 28814, 29071, 29164, 29249, 29580)
/gene="At2g31770"
<27609..29580
/note="At2g31770"
/join(27609..28814, 29071, 29164, 29249, 29580)
/gene="At2g31770"
/codon_start=1
/product="putative ARI-like RING zinc finger protein"
/protein_id="AAD32295.1"

Query Match      8.98; Score 183.6; DB 49; Length 99188;
Best Local Similarity 56.7%; Pred. No. 7,5e-27;
Matches 339; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY 804 atgacgacatgccctccgaatggttgagcagcaatccatcttcggtggttaacgtt 863
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15296 AAGCTAACGAATGCCCTTCATGCGTTCACCAAAACGCCGCGTTCAGTACATCAAGTGT 15395

QY 864 cgttcggaagcttcgttaatacagcagacgcgaatggaaggaagaagaagggcgt 923
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15356 CTTTGTGAGAGCTTGCCCGCTTTAAAGACATFCAATGATGAAGTGGCGCTGTCTAA 15415

QY 924 tagattgtggaagggccgttttctgtggttgtaagagtaaacgaaggaagaagatgta 983
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15416 AACCAATGAGGCCATTAACCTTCTTATGGGTGTGAGAACTGAAACAAAGAACTCCAA 15475

QY 984 taagtgcacgtgagaggttgaaacgaagtgagggaatatgcatcttggttccatattgg 1043
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15476 GCATTTACATAGAGACATTTGTGACAAAGGATTTGATGATGAGATTCCTCAATTAC 15535

QY 1044 aagtcctgacacatccctcgtttggagatgttcgtgcaacactgcggttgaaatcgacac 1103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15536 AAGTCTTGACATTAATCAATACGCTGTTTCATGACTCATGTCGGGGTGGAATTCGACTT 15595

QY 1104 tagaagatatactcttcggggttccgaatggttggtccttcgcagatggttcgataagga 1163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15596 TAGAGCATATGAGCTTACGAGTCTTGTGATGAGAAATGCCGCTTATACCGACCAAGCCA 15655

QY 1164 cgaatgcgaacactgataagagatgtctgtggaaggaacgggtgtggaagtgaaagctaaatgag 1223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15656 CTAATGCTAAGTTATTGAAGATGTGAGAGGTGGGTGAGGTAAAGCAAGATCAAA 15715

QY 1224 aaggttaagctcgttgatgagtgataaattcaggaagatgataaggggttataagatgggg 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15716 AAGGGTTTCTCCGAAGAGAGAGATGTGAGATGCTCTTGACAGAGTTAAGAAAGATAGT 15775

QY 1284 gagaagaaagatagagaacactgagaagatgtgagcaagtgaagaattcggaaagaaag 1343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15776 CGACAGAAAGGAGAGATAGTGAAGAAATATCCTCGAGAGTTGATGAGATTCGACAGGAAG 15835

QY 1344 ctatggaagaagatgagatctcaattcaaacactcaagtccttccttgatgaggtgtc 1401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15836 CTTTGTCTGATGAGAGAAATTCATAGAAATATGATGATGATTTGTTGCTAAATTTGT 15893

RESULT 14
AC005106/c
```

LOCUS	AC005106	84203 bp	DNA	PLN	16-DEC-1999
DEFINITION	Genomic sequence for Arabidopsis thaliana BAC T25N20 from				
ACCESSION	AC005106	Chromosome 1, complete sequence.			
VERSION	AC005106.2	GI:6587719			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 84203)				
AUTHORS	Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H., Conway, A., Conway, A., Kutz, D., Oji, O., Shen, Y., K., Toriumi, M., Vysotskaja, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and Ecker, J. R.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC T25N20 from				
JOURNAL	Chromosome 1				
REFERENCE	2 (bases 1 to 84203)				
AUTHORS	Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	3 (bases 1 to 84203)				
AUTHORS	Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	4 (bases 1 to 84203)				
AUTHORS	Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	5 (bases 1 to 84203)				
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Alatali, H., Bel, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzalez, R., Hansen, N., Howling, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharasy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thayer, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA				
COMMENT	On Dec 16, 1999 this sequence version replaced gi:3366536.				
FEATURES	Location/Qualifiers				
source	1..84203				
	/organism="Arabidopsis thaliana"				
	/db_xref="taxon:3702"				
	/chromosome="1"				
	/complement(join(1491..2159,2389..2452,2568..2703,3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))				
	/note="5' end of hypothetical protein YUP8H12.1; overlaps 7' end of YUP8H12"				
	/codon_start=1				
	/evidence=not experimental				
	/product="T25N20.1"				
	/protein_id="AAC80580.1"				
	/db_xref="GI:3935137"				
	/translation="MGKREKRDHHSVVEDSKVEAVLHLKHSPLTLKOEFCNRACVGRFLIKDQVKKAKQRLSCLSMRSLGIESILADEFTALAGLAVVAGLDECRFLVLRIRKODVKTOKQLTRVLFLEVALISIMRNEQVILFDALAAVLLVTLIVLSEFSLNVLILICNSTAHLPPFPFFLYKYNPKLQISIFHEHCKDTQNSRRGIRTFVDLSATMIYVSMQDPOSDPDSSSVRSLSRPTSSLSKMDKIGSAASSRACTFVSDDGLDYKPKCLITPSSSKRGHNGATISPLNKRSSSPASPARSPRCGRPSSTASTPMARKTDRHSIGTLRDPRIPRSPFOSPAIFRRESHSVSSSEKPRDSFVQELKTYRRPYDEMYRSMKPPGLGLVSVSTQIRRHVSLSQRF"				
CDS	join(9304..11268,11387..11513,11607..11692,12103..12341,12698..12784,13257..13356,13451..14263)				
	/note="unknown; similar to ESTs dbj C26311 and gb A1054650"				
	/codon_start=1				
	/evidence=not experimental				
	/product="T25N20.2"				
	/protein_id="AAC80581.1"				
	/db_xref="GI:3935138"				
	/translation="MEGGKSGEGRSRIASNTPSLKEKSDGVGRVYSTETRONOKRSMRVYDSESDDEMPPPRRSQVDRKTGAKKPYRRKDRVHEHNRNRYVRNNSAGSPMKNNKILDEPEDEYDPSANIMRRRPNNSGVGRSSSPASRRDYSVGRSSGRCLEPRRPNTPMTCSSASQSDSSSDSDPDRMGQINGYLKYNNKTKTLASTNPRDAIETRPSSRKAKRREANVYKPPFRKSNVNDNSSESDSKSRKKEISYKPKRENTKSTPELVNPDVBERGRGGTDQRRLREIRIKALMDAGTIDYKRRNDYDLAVYVNSGAYVSI1KAVDALLKQDKDGVAPARPKDTAAVASVEEIVNKLARAKKTRSEMRKKNKONSISKADGKATGDEGAYDTSEERIRSIKGGKSTKGRNGADWDELHKRSKSLYNNARPSGSDSHYLRGRTKRGCTLLVSSKDKKAPALNPNFYSGKRTPLSMILIESGVYOLRQVYMRRCAKVMEGMITREGIHCDCCSKILTVSREIHAGSKSCQPRONTIYESGASILQCCYRANMOKDAMNLALHOVDGDDPNDDAGIGGGDGLICDGCPSYTHONCLGMOYLPSSGDMPCNCTCKFCDAAVASGKDGNFISLSCGMCERRYHQLCLNDBAHKVVSFGSASFCCKPLIEPKLQYLYKTEIEGYSWSGLIHRVDTSDTNSQSAQRIEENMSKLAVGLIMDECTPIYDRSGVDLIRNVLNCGSNFRNLNRYGFTYLLERDETTISAASLRFHGMOLAMP1GTRHYTRQGMCRRLFPAISAMRSLKREKLV1PAP1PDLHAMTGKTEFTPLDSDVAREKMSLNTLVFPGIDMLQKPLHRENTIAPAAAGDAMISVEERKSEFTSSVEIGYAVAGDEVDVADAAKCYKIDILADSDENILVSVETAMGTICKRDELSHFPGESGSSPCQYTLKSGTKRHVLGHICDDPTGSCEDGLDVNVEADAPSLISOEIOQASASRVENINSLISGSGSSDLSSTSEVRSEKSSNLDGVPCXQOYNTILYPAKIDSKDDAPFADGFL"				
	/complement(join(14820..14957,15038..15420,15643..15721))				
	/note="hypothetical protein"				
	/codon_start=1				
	/evidence=not experimental				
	/product="T25N20.3"				
	/protein_id="AAC80582.1"				
	/db_xref="GI:3935139"				
	/translation="MGFLVAVNPSPTLVHNMKSPDOCNELRQGTSSLFDRGPKLCYVGAASSPATIEPFGSLQAQASEKLDGSEVCAFTLDPDRKRSKCELPKAYLKSANVVKIMRESLKENPKDNKAFKRSADKASITRDLISMMRKQKTVAGESEVELENIVIRALKESYKSPAPLPDEYKTEILDINKADEP"				
	/note="unknown; similar to EST emb1226799"				
	/codon_start=1				
	/evidence=not experimental				
	/product="T25N20.4"				
	/protein_id="AAC80583.1"				
	/db_xref="GI:3935140"				
	/translation="MMMDPPTEKFEFALAPYVYMDKRCQPVDDYDARVGCPIKRIILKRSNGPVTITAVGSLKVPRLLEVSGTGLSTHEVARPLEEEDSCSETPPASWICSVGRICGPGAGOGVNFITHLSTRHEHLEKRRFTNPGHLAN"				
	/complement(join(17577..18026,18231..18796,18981..19410))				
	/note="similar to RNA polymerase II elongation factor homolog (S59773)"				
	/codon_start=1				
	/evidence=not experimental				
	/product="T25N20.5"				
	/protein_id="AAC80584.1"				
	/db_xref="GI:3935141"				
	/translation="MFLHVVLIEMASIGFEVIECRSYNMDSKIDDSNIGSEPHLLARPVSLESGELPYAPENMNPMDVHMKVPRIGSGYFVDRYLTPPKYLPGLDTFLRKKVPRLSLQRYIRVHPREADQVQFAFSSGICSDGQDGLCCGERRCVDCCLISCKLISLEHGYSITKCAVYSEGHGCHVAMNACALRAYLACTIGSKGLDTETTCRCDAKNQDLPHYKFLIEICQVETQGDVETKLNGLGICLIGADQRNDAKELNCTESTVYKIIITLVVDVSDSGEARENDDTQSLQDVYPIGIPNHEAENKRLLEELGILRLRAOEFYQIEMERLHAQKECLSDYRQLEKESKLSISRVGTGANSILMTNVLKRDIOIRKEYTLLKMEAEVAGFGRTFGVLEEYFHLIND"				
	/note="unknown; similar to ESTs emb1233701 and emb1234709"				
	/codon_start=1				
	/evidence=not experimental				
	/product="T25N20.6"				

```

CDS
db 991136 protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translation="MNTATTSLTLEFLTKITPRVKKMFHFLCFPSNITKRSSEPGAT
SDDPKRPSTLLINFNFLIDDDSSAAHRRLSKPLIHDPVSSSTPASTVANSSSSA
SYSDSDGNFAPDDSDPPDLTAVALNRRFEESSGCGNSITPDSPLRCDNDYDTR
LLTGTAQKAVHYVOSPDPYNDNFRSMQEMDAITNAGDLRRREFLHELLLSYSLNAD
THKFIKRAFDILVSLSDGHRIS"
complement(join(26300..26503,26613..26880,27004..27259,
27394..27499,27748..27804))
/note="unknown: similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence-not-experimental
/product="725N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAAALNOLFVLASKPEOEKITPEESRAIVSCHFKALTAGFA
SVGGGLTWQYTKKKPKGLEVALNLAANYTHIHVYDHSRLSEKCKPFEFSTGO
ELLRLILIGIGVPMOHPPLISLWMLHECKRSLFVSNLNGEAMRQMLSKREY
PSVYGGDEGDKPMRMRRTTPELISYDPAKRNPNNGCIPNSHRISGSDA
SKRTQTONLNSGNDGEMAEVDIVIFCCSATESITAPVITSLAKTQTRQKRAQ
KRQRLKRNKASTNPQELA"
complement(join(28447..28740,28976..29125,29206..29443))
/note="hypothetical protein"
/codon_start=1
/evidence-not-experimental
/product="725N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translation="MKSRAENQMINRETKWCKLIVSPSEKSMSTSTTTTTTTTTVS
TEDMRSDNSCYFPDCKRDANCSCEICLDISLNLADLMLPSVOKSLJTKLSASNKS
LIVESPSTSPDTVVTTPAVASRPILKLMTSPKKLITKSKVFNEDRRTTKTERL
LIVLAKVILVGLVGLIELGFSWLEVEFKPEFEEIVRNAYRTOADLDLCKRL
LEDEKGEVETSRKFSRCTGSDSKKINOCPNLNCVLYSALEEYSINGMPLQVLEY
LEHTEFSSSTVLSGRVTEWTEGKFGYSIHETNLSMKTKWSTSVGLDPPNVTLEY
SOSVMDSSLSLSTIDIMEHLVPRVANNVRRNMFSSRLVNEAKSASMTPE"

CDS
join(32077..32407,32715..32955)
/note="similar to lipid transfer protein (AB007843)"
/codon_start=1
/evidence-not-experimental
/product="725N20.9"
/protein_id="AAC80588.1"
/db_xref="GI:3935145"
/translation="MNSNSFLISALIFSLSSNPTSILANOITPCSPSLSSVGC
TSLFGLGSPPTSCGALSKLTCGMDCLINAGPAGTIPINPLIASIPRACI
BOVPQCKGPASRFPPTTSPDSDTSPDETRRLPRTSSNSKTFCEPFSKRSGLL
TSHNSQGLSVRLVRLQLKLLLVNHNIS"
complement(33000..33155)
/note="similar to EST emb|Z26808"
/codon_start=1
/evidence-not-experimental
/product="725N20.10"
/protein_id="AAC80589.1"
/db_xref="GI:3935146"

Query Match 8.7%, Score 179.2; DB 49; Length 84203;
Best Local Similarity 51.5%; Pred. No. 5.7e-26;
Matches 740; Conservative 0; Mismatches 593; Indels 105; Gaps 10;

CDS
db 991136 protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translation="MNTATTSLTLEFLTKITPRVKKMFHFLCFPSNITKRSSEPGAT
SDDPKRPSTLLINFNFLIDDDSSAAHRRLSKPLIHDPVSSSTPASTVANSSSSA
SYSDSDGNFAPDDSDPPDLTAVALNRRFEESSGCGNSITPDSPLRCDNDYDTR
LLTGTAQKAVHYVOSPDPYNDNFRSMQEMDAITNAGDLRRREFLHELLLSYSLNAD
THKFIKRAFDILVSLSDGHRIS"
complement(join(26300..26503,26613..26880,27004..27259,
27394..27499,27748..27804))
/note="unknown: similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence-not-experimental
/product="725N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAAALNOLFVLASKPEOEKITPEESRAIVSCHFKALTAGFA
SVGGGLTWQYTKKKPKGLEVALNLAANYTHIHVYDHSRLSEKCKPFEFSTGO
ELLRLILIGIGVPMOHPPLISLWMLHECKRSLFVSNLNGEAMRQMLSKREY
PSVYGGDEGDKPMRMRRTTPELISYDPAKRNPNNGCIPNSHRISGSDA
SKRTQTONLNSGNDGEMAEVDIVIFCCSATESITAPVITSLAKTQTRQKRAQ
KRQRLKRNKASTNPQELA"
complement(join(28447..28740,28976..29125,29206..29443))
/note="hypothetical protein"
/codon_start=1
/evidence-not-experimental
/product="725N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translation="MKSRAENQMINRETKWCKLIVSPSEKSMSTSTTTTTTTTTVS
TEDMRSDNSCYFPDCKRDANCSCEICLDISLNLADLMLPSVOKSLJTKLSASNKS
LIVESPSTSPDTVVTTPAVASRPILKLMTSPKKLITKSKVFNEDRRTTKTERL
LIVLAKVILVGLVGLIELGFSWLEVEFKPEFEEIVRNAYRTOADLDLCKRL
LEDEKGEVETSRKFSRCTGSDSKKINOCPNLNCVLYSALEEYSINGMPLQVLEY
LEHTEFSSSTVLSGRVTEWTEGKFGYSIHETNLSMKTKWSTSVGLDPPNVTLEY
SOSVMDSSLSLSTIDIMEHLVPRVANNVRRNMFSSRLVNEAKSASMTPE"

CDS
db 991136 protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translation="MNTATTSLTLEFLTKITPRVKKMFHFLCFPSNITKRSSEPGAT
SDDPKRPSTLLINFNFLIDDDSSAAHRRLSKPLIHDPVSSSTPASTVANSSSSA
SYSDSDGNFAPDDSDPPDLTAVALNRRFEESSGCGNSITPDSPLRCDNDYDTR
LLTGTAQKAVHYVOSPDPYNDNFRSMQEMDAITNAGDLRRREFLHELLLSYSLNAD
THKFIKRAFDILVSLSDGHRIS"
complement(join(26300..26503,26613..26880,27004..27259,
27394..27499,27748..27804))
/note="unknown: similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence-not-experimental
/product="725N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAAALNOLFVLASKPEOEKITPEESRAIVSCHFKALTAGFA
SVGGGLTWQYTKKKPKGLEVALNLAANYTHIHVYDHSRLSEKCKPFEFSTGO
ELLRLILIGIGVPMOHPPLISLWMLHECKRSLFVSNLNGEAMRQMLSKREY
PSVYGGDEGDKPMRMRRTTPELISYDPAKRNPNNGCIPNSHRISGSDA
SKRTQTONLNSGNDGEMAEVDIVIFCCSATESITAPVITSLAKTQTRQKRAQ
KRQRLKRNKASTNPQELA"
complement(join(28447..28740,28976..29125,29206..29443))
/note="hypothetical protein"
/codon_start=1
/evidence-not-experimental
/product="725N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translation="MKSRAENQMINRETKWCKLIVSPSEKSMSTSTTTTTTTTTVS
TEDMRSDNSCYFPDCKRDANCSCEICLDISLNLADLMLPSVOKSLJTKLSASNKS
LIVESPSTSPDTVVTTPAVASRPILKLMTSPKKLITKSKVFNEDRRTTKTERL
LIVLAKVILVGLVGLIELGFSWLEVEFKPEFEEIVRNAYRTOADLDLCKRL
LEDEKGEVETSRKFSRCTGSDSKKINOCPNLNCVLYSALEEYSINGMPLQVLEY
LEHTEFSSSTVLSGRVTEWTEGKFGYSIHETNLSMKTKWSTSVGLDPPNVTLEY
SOSVMDSSLSLSTIDIMEHLVPRVANNVRRNMFSSRLVNEAKSASMTPE"

CDS
db 991136 protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translation="MNTATTSLTLEFLTKITPRVKKMFHFLCFPSNITKRSSEPGAT
SDDPKRPSTLLINFNFLIDDDSSAAHRRLSKPLIHDPVSSSTPASTVANSSSSA
SYSDSDGNFAPDDSDPPDLTAVALNRRFEESSGCGNSITPDSPLRCDNDYDTR
LLTGTAQKAVHYVOSPDPYNDNFRSMQEMDAITNAGDLRRREFLHELLLSYSLNAD
THKFIKRAFDILVSLSDGHRIS"
complement(join(26300..26503,26613..26880,27004..27259,
27394..27499,27748..27804))
/note="unknown: similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence-not-experimental
/product="725N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAAALNOLFVLASKPEOEKITPEESRAIVSCHFKALTAGFA
SVGGGLTWQYTKKKPKGLEVALNLAANYTHIHVYDHSRLSEKCKPFEFSTGO
ELLRLILIGIGVPMOHPPLISLWMLHECKRSLFVSNLNGEAMRQMLSKREY
PSVYGGDEGDKPMRMRRTTPELISYDPAKRNPNNGCIPNSHRISGSDA
SKRTQTONLNSGNDGEMAEVDIVIFCCSATESITAPVITSLAKTQTRQKRAQ
KRQRLKRNKASTNPQELA"
complement(join(28447..28740,28976..29125,29206..29443))
/note="hypothetical protein"
/codon_start=1
/evidence-not-experimental
/product="725N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translation="MKSRAENQMINRETKWCKLIVSPSEKSMSTSTTTTTTTTTVS
TEDMRSDNSCYFPDCKRDANCSCEICLDISLNLADLMLPSVOKSLJTKLSASNKS
LIVESPSTSPDTVVTTPAVASRPILKLMTSPKKLITKSKVFNEDRRTTKTERL
LIVLAKVILVGLVGLIELGFSWLEVEFKPEFEEIVRNAYRTOADLDLCKRL
LEDEKGEVETSRKFSRCTGSDSKKINOCPNLNCVLYSALEEYSINGMPLQVLEY
LEHTEFSSSTVLSGRVTEWTEGKFGYSIHETNLSMKTKWSTSVGLDPPNVTLEY
SOSVMDSSLSLSTIDIMEHLVPRVANNVRRNMFSSRLVNEAKSASMTPE"

CDS
db 991136 protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translation="MNTATTSLTLEFLTKITPRVKKMFHFLCFPSNITKRSSEPGAT
SDDPKRPSTLLINFNFLIDDDSSAAHRRLSKPLIHDPVSSSTPASTVANSSSSA
SYSDSDGNFAPDDSDPPDLTAVALNRRFEESSGCGNSITPDSPLRCDNDYDTR
LLTGTAQKAVHYVOSPDPYNDNFRSMQEMDAITNAGDLRRREFLHELLLSYSLNAD
THKFIKRAFDILVSLSDGHRIS"
complement(join(26300..26503,26613..26880,27004..27259,
27394..27499,27748..27804))
/note="unknown: similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence-not-experimental
/product="725N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAAALNOLFVLASKPEOEKITPEESRAIVSCHFKALTAGFA
SVGGGLTWQYTKKKPKGLEVALNLAANYTHIHVYDHSRLSEKCKPFEFSTGO
ELLRLILIGIGVPMOHPPLISLWMLHECKRSLFVSNLNGEAMRQMLSKREY
PSVYGGDEGDKPMRMRRTTPELISYDPAKRNPNNGCIPNSHRISGSDA
SKRTQTONLNSGNDGEMAEVDIVIFCCSATESITAPVITSLAKTQTRQKRAQ
KRQRLKRNKASTNPQELA"
complement(join(28447..28740,28976..29125,29206..
```

[illegible]

```

0Y 1287 aaagagtaggaagaacttagtagagagagcgcgagcgaagtcggaagatttcgcaagaagaagctca 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77916 AGAACTCGGCTGGAATTGAGGAAACGCAAAAGAAATGAGACGCTTTAGCCGTGGAACCGG 77857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 1347 tgaaggaagatgcatcttcagtttaacaactcaagctcttcctgcatgaggttcgtag 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77856 GTAGAGAAAGAGGATCTTCGATTAAGAACATGACGCTTTGTGGAGAGATATTTGTGCG 77799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
A62523
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1. 1513
/organism="unidentified"
/db_xref="taxon:32644"
69..1469

```

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAA03659.1"
/db_xref="GI:3716431"
/translation="MAETTTTATNSKGAVVLPYVOGHLNPMVOAKRLVSKGVN
VITATTPYPASSISTPVSVEPIDGDHETILIGPGVSIIDAYESKRLNGSTELRLVY
SKRTSTDSIPDLVYDSELPWGLLEVARNSISAAAEFTNNLWCVSLRRFVGEPLLE
ADAPSAPLYVRLGTPALSTYDELPSFVGRHSSHSDKGVGLNLRNHEADMLFVSPFS
GLETCQCEYGESEANRATLIGPMIPSAVLDGRIKDDKGVGLSSLMKRLSECEHEWLDIT
LSKSVFVSFFSGFGLLEFKQLAEYAKALQDSNRPFLMTVLEKHAIALPEGFVEAKRDK
ALLVSCNQLLEVLAHVSIGCFLTICGNNSTLEGISLGVPMVGPVQSDMDNARKVEEY
YMHVYRAKEAGGCGVVCSDVEVRCILRGVMEGSGSEVIRRESSKRWMDLAVKAMSGGS
SDRSINLPESLSGRKH"
BASE COUNT      368 a      356 c      399 g      350 t
ORIGIN

```

Query Match	8.7%;	Score 178.4;	DB 5;	Length 1513;
Best Local Similarity	53.0%;	Pred. No. 7.9e-26;		
Matches 405;	Conservative	0;	Mismatches 356;	Indels 3;
				Gaps 1;

QY 659 gaagaaaaaaccttaagagcttcgtggaacacagctttgcttgcttggtggagccctgaatgcctaaag 668

Db 667 GAAGATGCTGATTTGGCTGCTTCGCAACACCTTCGAAAGGCTTAGAGACCAAGGCTTGAA 746

QY 689 gccatgtgaagtccgaatgcatgtgaacacgagccgttgcattccctccgaccttcctgagc 748

Db 747 GTTGAGAGATCAGAGCGGATGAGGCGGACGCTGATTCGCGACATGATATCCATCTGGTTAT 806

QY 749 ggtlaaagatcccttcggagcagctctcttcgscgagatctgttcggaaagagtgctgaagac 808

Db 807 CTCGACGCGCCGAATCAAGACGATTAAGGCTACGGCTCGAGCCGATGAAAGCCGCTTCG 866

QY 809 gaacattgcctcgaaatggttgtagcagaatcctgcatacttcgtggtgtttagcgtttc 868

Db 867 GAGGAGTATAGAGATGATGATGATTAACACTTAAGTCGAGCAAGTCGATGTTTGTGTCATT 926

QY 869 gaaagcttcgttlaatacagcagaatctgcgcaaatgaaagataagcaagaagctgtttagat 928

Db 927 GCTTCCTTTTGGGATCCTCTTTTGAGAGCAACTCGCTGAGGTGCGCAAGCGCTTACAAGAA 986

OY	929	tgtgaggagccgtttcttctgtggtgtaaggaatgaagaagaagagatttgtaagt	988
Db	987	TCGAACCTTTAACTTCTCTGTGGGTATCAAAAGAGCTCATATAGCGAAAGTTACCGAAGG	1046
OY	989	tgcacgaggagttcgaacgaatgagggaanaattgtactctgtgtctctcaattggaagtc	1048
Db	1047	TTTGTGTGAAGCTACCAAAAGACAGAGCCCTCTCTGTTCTTGGTGTAAACCGACTGTGAGATT	1100
OY	1049	cttagccatccctctgttggagtgtttctgtgaacacctcggtggatgaatttgcattcaag	1100
Db	1107	TTTAGCTCATGTTATCATATAGTGTGCTTTTATACACACTGGGCTGTGAACCTGCAAGTTGGAA	1166
OY	1109	agatactcttcgggggttcceatgtgtgctcttcocgaatgcttcgaataaggaacaaat	1168
Db	1167	GGATTGAGTTTGGAGATTCGCATGCTGTGTGTGTCCCTCACTGAGAGATGATCAAGATGAT	1222
OY	1169	gcgaacctcgaatgaggaatctgttggagaacgggtgtgaaagtgtgaagcttaattgagaaggt	1222
Db	1227	GCTAACTTTGTGTGAAGAGCTTTTGGAGCTTGGGTATATAGCGAAGGAGGAAGCTGGGGGA	1286
OY	1229	agagctgttgatgagtgatgaattagaagaatgattatgtaggggttatgagatggggagaaga	1288
Db	1287	GGAGTTCTGAACAGCGATAGAGCTGTGACGTGTTTGACAGCATGTGATGA---AGGAGAG	1344
OY	1289	aagaatgaggaactatgagagagatctctcgtgaagaatggaagaatttggcgaagaagaagcat	1348
Db	1344	AGTAGTGTGAATATTAGAGAGAGATTTCATCAACATGTGAAGAATTTGGCTGTGAAGCGGATG	1403
OY	1349	gaggaagatgagttctcaagtaaacacactcaaggtctttcttga	1392
Db	1404	AGTGAAAGAGAGAGCTCTGATCGGAGGATTAAAGATTTATGTGA	1447

Search completed: August 1, 2000, 13:07:11
Job time: 55893 sec

1990

3

THIS PAGE BLANK (USPTO)

Query Match	100.0%	Score 2062;	DB 1;	Length 2062;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2062; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0

QY	61	agcaacgggagacacataaatccgcgccttcaatttcgcgaagcgcttcgcgaatagccgaat	120
Db	61	AGCACAGGGGACATATAATCCGCCGCTTCAATTGCGCAAGCGCTTCGCATAAGCCGACAT	120
QY	121	tcaagtcacatctctaccagcgctcacaagctgagcgccgatctccagaaccgcgcgtg	180
Db	121	TCAAGTCACATTTCTTACCACAGCGCTTACGCATGCGCGCGCATGTCAGAACCGCGCGTGG	180
QY	181	ctcaaacggcgctcacaattctgtgctgtttccgacgggtabgacagcggtttacagcc	240
Db	181	CTCAAAAGGGCGTCATCAATTTGTGTGCTTTCCGACGGGTATGACGACGAGGTTACAGCC	240
QY	241	cggagacgattggaagaactacatgctcgugagtgaaagaagagagatataaaagccttgag	300
Db	241	CGGAGACGATGGGAAGACTCATGTCGGGAATGTAAAGCAGAGGATATAAAAGCCCTTGAG	300

QY 301 cgatactcttgagccaataatgctgatacaaaaagcagcaaaatccacgtctcggtgta 360
 Db 301 CGATACTCTTGAGCCCAATAATGCTGATCAAAAAAGCAGCAAAATCACGTTCTGCTGTGA 360
 QY 361 ctcccaacctcttgatgagcgccgaaggtgctgctgagttccatctccgagcgct 420
 Db 361 CTCCCAACCTCTTGATGAGCGGCCGAAGGTGCGGTGATTCATCTCCGAGCGGCT 420
 QY 421 actctgagatgagcagcactacggtgtgtgatatattactcttattccaacgctatag 480
 Db 421 ACTCTGAGATGAGCAGCAGCTACGGTGTGATATATTTTCTTTTCAAGGCTATAG 480
 QY 481 cgaacgaacacgacgaggtccgagctatccatctcccgagagagaccacagtcgtgc 540
 Db 481 CGACGAATCGATCGGGTTCGGATGCTATTCATTGCCGAGAGACCTCCAGTCTGTC 540
 QY 541 ccagcgatattacgcgtctctctctctctccacgcatgagagatccgtctacat 600
 Db 541 CCAGCGATATTACGCGTCTCTCTCTCTCCACGCAATGAGAAATCCGCTTCATGAT 600
 QY 601 gaaggaagaattggaaccttgaagaagtgaagaaaaaacccaagtccttgatgaacagct 660
 Db 601 GAAGGAGAAATTGGAACCTTTAGAGAGTGAGAAAAAAGCTTAAGCTTGCTAACAGCTT 660
 QY 661 tgaagcgttgagacgtgacgtcgaagcgaatgataagtaacgagatgataatcg 720
 Db 661 TGATGCGGTGAGGCTGATGCGCTCAAGGCCATTTGATTAAGATGATTTGCAATGCG 720
 QY 721 gccgttga ctccctccgcatctctgagcagtaaaagatccctcgaagctcttcgagcg 780
 Db 721 GCCGTTGATTCCTTCCGATTTCTTGAGCGGTAAGAATCCTTGAGACAGGTCTTTCGGGG 780
 QY 781 agatctgtctgagaagaaggtcgaaatgagacgaatgctccgaatgctgagagacgaatcc 840
 Db 781 AGATTGTCTGAGAAAGGTCGAATGACGAGATTCCTCGAATGCTTGAACAGCAATCC 840
 QY 841 tccatctcgttgagcttaagctctcgttcggaagcctcgtttaa taagcagcagctcgaat 900
 Db 841 TCGATCTTGGTGTACGTTGCTTGGTTCGAGACCTTGTTAATACAGCAATCCGAAT 900
 QY 901 ggaagagaatgaaagagcgctgtaagcttgagagcgctcttctgtgagtgtaagct 960
 Db 901 GGAAAGATACCAAGAGCGCTGTATGATGTGGAGCGCGTTTGTGGCTGTAAAGCT 960
 QY 961 aaacgaagaagaagatataatgataatgacatgagagagctgaaacgaagtgagagaaat 1020
 Db 961 AAACGAAGAAGAAGATATGATTAAGTATGATGAGAGATTGAACGAGTGGGAAT 1020
 QY 1021 tgaatctgagctctcaatctgaaagcctgagcgaatccctcgttgagagtgcttgatc 1080
 Db 1021 TGAATCTGAGCTCTCAATCTGAAAGCCTGACGCAATCCCTCGTGGGCTGTCTGATC 1080
 QY 1081 acaatgagagtgagaaatctgaaactcagaagatatactctcggagctccgagtgagctt 1140
 Db 1081 ACACGCGGAGGAGATTCGACTACAGAGATATCTTCCGGGATCCGATGTGCTTT 1140
 QY 1141 tccgaagtgagctgaaatgaaagagcgaatgagagcgaatgagagagtgagagagag 1200
 Db 1141 TCCGAGTGCTTCATCAAGAGGAGATGCGAATGCGAATGAGAGAGTGTGAGAGCGG 1200
 QY 1201 tgaagagtgagagcctaagagagagtgagcgtcgttgaatgagtgagagagagag 1260
 Db 1201 TGTGAGAGTGAGAGCTAATGAGAGGCTAGCGCTGTGATGATGAATTAAGAGATG 1260
 QY 1261 tatgagagagtgatgagatgagagagagagagagagagagagagagagagagag 1320
 Db 1261 TATTGAGAGGTTATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1321 gtggaagagatctgagcaagaagaagctatgagagagagagagagagagagagagag 1380
 Db 1321 GTGGAAGAGATTGCAAGAAAGAGCTATGAGAGAGATGATCTTCAAGTAACACTCAA 1380

QY 1381 ggtctcttgatgaggttgatgatactaaagcgtaaatgaggtcccatagagcaaa 1440
 Db 1381 GGTCTTCTTATGAGGTGATGATATCTAAAGCAGTAATGAGGTCCCATAGGCAAAA 1440
 QY 1441 tgcgaatctcatctcgaagtgaaatctcttgagcttgaatttctgagagctt 1500
 Db 1441 TTGCAAAATTCATCTCGAAGTGAATACCTTTTGCTTTAATTTGTTGAGTTGTT 1500
 QY 1501 ttcgaatctcatctctgtaattctcaatgagtgtaaatctagctcgaatcttaacgaaa 1560
 Db 1501 TTCAAAAATTTATCTTGATATTTTATCAATGAGTGAATTAATTTAGCTGATTTAATCGAAAA 1560
 QY 1561 aatataaatctcatcttgagagcctcacaacaacatcatctgattccttaattgcttg 1620
 Db 1561 AATATAAATTCATCTTGAGAGCTCTTATCAAAATCAATGATTTCCCTTATGCTTGG 1620
 QY 1621 gtcaaaatctcatcaatctgagaaataaataatctcaaaatcgtccaaatcttgaa 1680
 Db 1621 GTCAAAATTCATATCAATTTGAAAAAATTAATTTCAAAATCGTCAATTTTGAACCA 1680
 QY 1681 gaaagaagataatcttgacaaataaataaagatccaagtgatcctgataagtgct 1740
 Db 1681 GAAAGAATATAATTTGACCAAAATATATAAGATTTCAAGTATGATGATGATGCT 1740
 QY 1741 gagcgaagatctatattctccacggaatctcgaagcttctgaatcttcttgagc 1800
 Db 1741 GAGCGAGAGTTCTATATTTTCCACGAATTTCTAACAGATTTTGAATTTTGTAGC 1800
 QY 1801 caaaatcgaactaaacttgtaaaaatgaaagctatagatgaatcttaaaaaacaa 1860
 Db 1801 CAAAATCGAGACTACTTTGTACAAAATGAAAATTTATATGATGAATTTTAAAAACAA 1860
 QY 1861 ctcaagaacataataagcccgaaagtgataaaatctccgaagaaatctgcaattcgcc 1920
 Db 1861 CTCAAGACATATATAAGCCGAAAGTGTAAATTTACCTGACCAAAATTTGCAATTTGCC 1920
 QY 1921 tccatcttaattcttctgagtgcttaataaaatcggcttaattcttaataaa 1980
 Db 1921 TCCATATTTAATTTTGTGCTTAAATTAATCGGTTATTTACTTTAATTTAAATA 1980
 QY 1981 aagtgagatgcataagatcgttgatgagatataatgagtgagtgagtgagtgatct 2040
 Db 1981 AAGTGAGATGCATGATGCTTGCTGATATATATGATGATGATGATGATGATGATGAT 2040
 QY 2041 tctaaaaaataaaaaaataa 2062
 Db 2041 TCTAAAAAATAAAAAAATAA 2062

RESULT 2
 X02826 standard; cDNA; 1506 BP.
 AC X02826;
 DT 14-MAY-1999 (first entry)
 DE W0905287 Seq ID 1.
 KW Plant; flavonoid 5-transglycosylation activity; 5Tc; variety; colour; ds.
 OS Perilla frutescens.
 FH key
 FT location/Qualifiers
 FT 17..1399
 FT /*tag= a
 FT /product= "protein with flavonoid 5-transglycosylation activity"
 FT
 PN W0905287-A1.
 PD 04-FEB-1999.
 PR 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNK) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 PI Yamazaki M,
 DR WPI: 99-142940/12.
 DR P-PSDB: W92947.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 PT used to transform plants for improvement of plant coloration

Query Match 33.08; Score 680; DB 1; Length 1474;
 Best Local Similarity 70.58; Pred. No. 2,56-145;
 Matches 1019; Conservative 0; Mismatches 370; Indels 57; Gaps 6;

9 aaaaaataaaaaaaatgagcagagctcagctcctctctgagcattccagacag 68
 12 ACAAATTTCCACAAAATGTTCCGCCGCCGCTGCTGTACCAACGTTTCCGGCGCAG 71
 69 gacacataatccgcgcctcattccagcagcgtctcgaatgacagcattcaatca 128
 72 GCCACATTAATCCGCCCTCCAAATTCGCCAAGAGACTCTTAACCCGCCAGTCGCA 131
 129 catctaccacgctcagcagctgagcagcagcagcagcagcagcagcagcagc 188
 132 CGTTTTCACAGCGTTTATGATGAGCGGCCCATGCGCAACACGCTCCGCCGTGCG 191
 189 g-----gctcataatttctgtcttctcgcagcagcagcagcagcagcagc 239
 192 GAACCCACCGCGGCTCGACTGCTGCTCTCCAGCGGTCAGACAGCGGCTGAAGC 251
 240 cggagagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 299
 252 CCGGGGGGAGGAGCGGACCGCTACATGCGAGATGAACCGCGGCTCCGAGCGCTTA 311
 300 cggatacctctgcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
 312 G-----AACTCTCTCTCAACAGACAGCGCTCTTCTGCTCT 353
 360 actccacactctctgcaatgagcagcagcagcagcagcagcagcagcagcagc 419
 354 ACTCCACCTCTTGTGATGAGCGCGGAGGCGGCTGCTGCTCCACGCGGACCGCC 413
 420 tactctgagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 479
 414 TTCTCTGGGTGAGACCGCCGACCGCTGCTGCTATACATTTCTTCAACGGCTACG 473
 480 ggcagaaatcagatgagcagcagcagcagcagcagcagcagcagcagcagc 539
 474 CAGACGAGATGAGCGCGGCTTCCAAATTAATCA--GCTCCCTGGCTTCATCCCTG 530
 540 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 599
 531 AGCAGCGAGCTCTTCCGACGCTTCTGCTGCTGCGAGCGCGGAGATTCGGTTGATGA 590
 600 tgaagagaaattgaaactttagaagtgaaagaaaaaactaaggctctgtgtgaagc 659
 591 TCAAGGAGAACTGAAACTTATAGACGCTGCAAGAGCGGAAAGATTTGGGAAACGCT 650
 660 ttgatgagcttgagcagcagcagcagcagcagcagcagcagcagcagcagc 719
 651 TTGATGCTTGAGACCGGATGACTCGGCTATTTAGTATGATGATGATGATGATG 710
 720 ggcgcttctctctcagcagcagcagcagcagcagcagcagcagcagcagcagc 779
 711 GGCCGTTGATTCCTCCGCTTCTTGGACGCGCAAGATCCGCGCAACGCTTACGCG 770
 780 gagatttcttgagaaagagtgcaatgagcagcagcagcagcagcagcagcagc 839
 771 GCGATCTTTTCAAAA--TCGAGAGAAATACGCTGAGAGGCTTGAACCTGGAAGC 827
 840 ctgcagctctggtgtttagcgttctgtctggaagcctgttaatacgaagagtcgaa 899
 828 CGAAATCTTCGCTGCTGATGCTGCTTGGAGCGTTTTCAGGTTTCCAAAGCAAAA 887
 900 tggagagagatgcaagagagcgttagatgtgagcagcagcagcagcagcagcagc 958
 888 TGGAGAGATTTGGAGAGGCTATTAACCTGCGGAAGGCCCTTTTATGATGATGATG 947
 959 -----gtaaacgaagagagagaglatatgaatgttcagatgagcag 998
 948 AACAGAAATGACGAGCGGAGAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAG 1007

999 agtgaacagcagctgagcagcagcagcagcagcagcagcagcagcagcagc 1058
 1008 AATTCAAAAAATGCGGAAAAATGCTGCGGCTCCGACGTTGAGCTGTGCGCCACC 1067
 1059 cctcgttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1118
 1068 CTGCGCTTGAGATGTTTGTGACGATTTGTGCGAGAACCTCGGCTGTGAGACCTT 1127
 1119 tcggagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1178
 1128 GCGGATTCGCGTGTGCTGCGGCTGCGGAGTGTGTTATGACGACGAGATGGAAGCTGA 1187
 1179 tgaagatctgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 1238
 1188 TTGAGAGATGCTGCGGAGCAAGGCTGAGATGAGATGAGATGAGAGCGGT---GGG 1244
 1239 atgtgtatgaaatgaagagatgcatgagcagcagcagcagcagcagcagcagc 1298
 1245 ATGATGTGAGATGAAAGCTGTGCTGAGATGCTGATGATGAGGAGTGAACAAC 1304
 1299 aacttaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1358
 1305 TAGTGAGAGAAATGCGATCAATGAGAGACTTGGCCAGACGACCATGGGATGATG 1364
 1359 gatcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1418
 1365 GATCTTCACTCAACATCTCAACGCTTCTTCTGTCAGATGTCACATTTAATCTGCTC 1424
 1419 aatgag 1424
 1425 AATCAG 1430

RESULT 4
 X02829
 ID X02829 standard; cDNA; 1671 BP.
 AC X02829;
 DT 14-MAY-1999 (first entry)
 DE W0905287 Seq ID 4.
 KW Plant; flavonoid 5-transglycosylation activity; 5mg; variety; colour; ds.
 OS Torontia hybrida.
 FH Key Location/Qualifiers
 FT CDS 45..1481
 FT /tag= a
 FT /product= "protein with flavonoid 5-transglycosylation activity"
 FT
 FT W0905287-A1.
 PD 04-FEB-1999.
 PR 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNR) SUNTORI LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y, Yamazaki M, WPI; 99-142940/12.
 DR P-PDB; M92950.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
 PS Disclosure: Page 60-64; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut CC flowers harvested from them.
 CC Sequence 1671 BP; 530 A; 344 C; 399 G; 396 T;

Query Match 23.58; Score 484.4; DB 1; Length 1671;
 Best Local Similarity 63.48; Pred. No. 5,9e-101;
 Matches 920; Conservative 2; Mismatches 453; Indels 75; Gaps 9;

9 aaaaaataaaaaaaatgagcagcagcagcagcagcagcagcagcagcagcagc 68
 31 AATTAATAAAAAAATGTTAACAAACGCAATTTCTACTAGAGAAATTCACAGCAAG 90


```

QY 191 cccatcatttctgttgcgttccgacgggtatgaagcggttcaagccggagagat 250
    |||||
Db 509 GGATTGAATTCATTCATTCATTCATTCGATGGCTTTCATGAACGTTTATCATTCAAAGAC 568
QY 251 gggaaagaaactacatctcggagatgaaaaagacagagataaaagccttgcatacctt 310
    |||||
Db 569 CCTGATTATTACATCTCACACACTTCGTAATGTGGAAAGTGAACCTGCAAAAAAATAATT 628
QY 311 gcaagcaataatgctgcacaaaagaacagcaaaatcgcgttcgtgtgctaccaccactc 370
    |||||
Db 629 -----CTCAGTCTGCTGAAAATGAGACAGCTTAACCTTGCTTACTTTCATTCATTTTC 682
QY 371 ttgcacatggcgccaaagtgtagcggttgcattccatctccgagcgctactctgatt 430
    |||||
Db 683 CTTCCTTGGGAGCAGAGTAGCACGTGAAGTTCCATCCCTTCTGCTCTTGGAGGT 742
QY 431 gggcagcctacgggtgttgatataatttacttatttcaagcgctatagcgaatc 490
    |||||
Db 743 CAACCGACACAAATATGACATATATTAATTAATTAATTAATTAATTAATTAATTAATG 802
QY 491 gatcgcggttcgagctatcacttccccc-----ggaaggaatcccaatgctgagc 541
    |||||
Db 803 GCGAATGAATCCAAATGATCAAAATTTGGTCAATTCATTCGCGGCTTCACACTAGCGAA 862
QY 542 cagggtgatcttcgcttcttccctcctcct-----ccagcagatgagagatc 589
    |||||
Db 863 ACTCGAGATCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 922
QY 590 cgttcacatgataagagaataatgaaacttgaagtgtaagaaacctaagctctg 649
    |||||
Db 923 CTTCACACATTCAAAGATTTGATGACATATGATGATGATGATGATGATGATGATGATGAT 982
QY 650 gggaaagccttgatgctgttgagcctgagctgagctcgaagccttgaatgaagatg 709
    |||||
Db 983 GCTGATATCATTTTGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1042
QY 710 atgcacatcgagcgcttgccttgccttgccttgccttgccttgccttgccttgccttgc 769
    |||||
Db 1043 TATGGAATTTGACACCTTATTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1102
QY 770 tcttcgagggagatttgcgagaaaggtgcgaatgacagatgcctgcgaatggttg 829
    |||||
Db 1103 TCAATTTGGTGATCTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1153
QY 830 agcagcgaatcctcgaatcctcgtgtgacgcttgccttgccttgccttgccttgccttgc 889
    |||||
Db 1154 AACTCAAAACCAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1213
QY 890 aagtcgcaaatgaaagataagcaagagggcgttgaattgtgagagcgcttcttggtg 949
    |||||
Db 1214 ATTAGCAAAATGAGAGAAATATCAAAAGGTTGATTAACATTAAGAAAGCCGCTTTTATG 1273
QY 950 gtgttgaaagata-----aacgaaggaagaaaggtatgtgaagttgcaatgagag 997
    |||||
Db 1274 GTGATATAAAGAAATGAAAGGCAAAAGAGAAAGAAAGAAAGAAAGCTTGTTGATTCGA 1333
QY 998 gagttgaagaaaggtgaggaagaaatgtatcgtgtgtcctaattgaaagcccgagcgct 1057
    |||||
Db 1334 GAATTTGAAAAAATAGAAAAATAGTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1393
QY 1058 cccctcgttgagatgttgcgtacacacatgctgagtgagaaatcgaactaagagatlatc 1117
    |||||
Db 1394 CCATCTTTTGTAGTGTGTTTGTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1453
QY 1118 ttcgaggttcgagatgtgttcttcgacaggtgttcgataagagcgaaatgcaagcgctg 1177
    |||||
Db 1454 TGTGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1513
QY 1178 acgagagatgttgcgagacggtgtgcgagagctgaatgagagagagctgctcgtc 1237
    |||||
Db 1514 GTTGAAGATGTGTGAAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1570

```

```

QY 1238 gatgtgatgaataatagagatgatatgagaggtatgagatggggagaaaaagatag 1297
    |||||
Db 1571 GAAGGTGAGAGAAATCAAAAGGTGATTAATTCATTAATTCATTAATTCATTAATTCATTA 1630
QY 1298 aaactbagaagagctgcgcgaagtggagagatttgcgaagaaacctaaggaagat 1357
    |||||
Db 1631 GAATTCAGAAAGAAATGCTAAGAAATGGAAGAAATTTGGCTAGACACCTGTGAAGGAGCT 1690
QY 1358 ggaattcttaagtaaacactcaaggtcttcttgatgaggttg 1400
    |||||
Db 1691 GGATCTTCACACAAATTTAAAGCCTTTTATTCATGATGATG 1733

RESULT 6
ID T66166 standard; cDNA to mRNA; 1513 BP.
AC T66166:
DT 15-JUL-1997 (first entry)
DE UDP-glucose:thiohydrotimate S-glucosyltransferase cDNA clone pGL9.
KW Glucosinolate; UDP-glucose:thiohydrotimate S-glucosyltransferase;
KM S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
OS Brassica napus cv. Westar.
FH Key
FT cds
FT 69..1469
    location/Qualifiers
    /tag= a
    /tag= b
    /note= "C is T in clone pGL6-14; causing an amino
    acid change from Ala-2 to Val-2"
    /tag= c
    /note= "G is A in clone pGL6-14; there is no amino
    acid change"
    /tag= d
    /note= "codon AAG (Lys) is present between clone
    pGL9 bases 98 and 99 in clone pGL6-14"
    /tag= e
    /note= "G is C in clones pGL6-14 and pGL3-22.
    causing an amino acid change from Val-43
    to Leu-43"
    /tag= f
    /note= "T is C in clone pGL3-22; there is no amino
    acid change"
    /tag= g
    /note= "T is C in clone pGL3-22; causing an amino
    acid change from Leu-75 to Pro-75"
    /tag= h
    /note= "A is C in clone pGL3-22; there is no amino
    acid change"
    /tag= i
    /note= "A is G in clone pGL3-22; causing an amino
    acid change from Glu-88 to Gly-88"
    /tag= j
    /note= "A is C in clone pGL3-22; causing an amino
    acid change from Asn-93 to His-93"
    /tag= k
    /note= "G is C in clone pGL3-22; causing an amino
    acid change from Glu-96 to Gln-96"
    /tag= l
    /note= "A is C in clone pGL3-22; there is no amino
    acid change"
    /tag= m
    /note= "A is C in clone pGL3-22; there is no amino

```

FT acid change"
FT 398
FT /tag- n
FT /note- "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 410
FT /tag- o
FT /note- "A is T in clone pGL3-22; there is no amino
FT acid change"
FT 458
FT /tag- p
FT /note- "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 465
FT /tag- q
FT /note- "A is C in clone pGL3-22, causing an amino
FT acid change from Ile-133 to Leu-133"
FT 473
FT /tag- r
FT /note- "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 512
FT /tag- s
FT /note- "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 515
FT /tag- t
FT /note- "A is T in clone pGL3-22; there is no amino
FT acid change"
FT 526
FT /tag- u
FT /note- "T is C in clone pGL3-22, causing an amino
FT acid change from Ala-153 to Val-153"
FT 568
FT /tag- v
FT /note- "C is T in clone pGL3-22, causing an amino
FT acid change from Pro-167 to Leu-167"
FT 587
FT /tag- w
FT /note- "A is G in clone pGL3-22; there is no amino
FT acid change"
FT 623
FT /tag- x
FT /note- "C is G in clone pGL3-22; there is no amino
FT acid change"
FT 647
FT /tag- y
FT /note- "G is C in clone pGL3-22; there is no amino
FT acid change"
FT 678.
FT 679
FT /tag- z
FT /note- "bases 678-679 are AT in clone pGL3-22,
FT causing an amino acid change from Arg-204
FT to Ile-204"
FT 685
FT /tag- aa
FT /note- "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 714
FT /tag- ab
FT /note- "A is G in clones pGL3-22 and pGL4-2,
FT causing an amino acid change from Ser-216
FT to Gly-216"
FT 759
FT /tag- ac
FT /note- "G is A in clone pGL3-22, causing an amino
FT acid change from Ala-232 to Thr-232"
FT 761
FT /tag- ad
FT /note- "G is C in clone pGL3-22, causing an amino
FT acid change from Ala-232 to Thr-232"
FT 766
FT /tag- ae

FT /note- "G is A in clones pGL3-22 and pGL4-2,
FT causing an amino acid change from
FT Arg-234 to Lys-234"
FT 794
FT /tag- af
FT /note- "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 809
FT /tag- ag
FT /note- "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 814
FT /tag- ah
FT /note- "G is C in clone pGL3-22, causing an amino
FT acid change from Gly-249 to Ala-249"
FT 848
FT /tag- ai
FT /note- "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 901
FT /tag- aj
FT /note- "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 936
FT /tag- ak
FT /note- "G is A in clone pGL3-22, causing an amino
FT acid change from Gly-290 to Arg-290"
FT 959
FT /tag- al
FT /note- "C is A in clone pGL3-22; there is no amino
FT acid change"
FT 968
FT /tag- am
FT /note- "G is A in clone pGL3-22; there is no amino
FT acid change"
FT 973
FT /tag- an
FT /note- "A is C in clone pGL3-22, causing an amino
FT acid change from Lys-302 to Thr-302"
FT 1013
FT /tag- ao
FT /note- "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 1117
FT /tag- ap
FT /note- "G is T in clones pGL2-7 and pGL2-25,
FT causing an amino acid change from
FT Val-350 to Glu-350"
FT 1121
FT /tag- ap
FT /note- "G is T in clone pGL2-7; there is no amino
FT acid change"
FT 1130
FT /tag- ar
FT /note- "C is T in clone pGL2-7; there is no amino
FT acid change"
FT 1161
FT /tag- as
FT /note- "T is C in clone pGL2-7; there is no amino
FT acid change"
FT 1169
FT /tag- at
FT /note- "A is G in clone pGL2-7; there is no amino
FT acid change"
FT 1193
FT /tag- au
FT /note- "T is G in clone pGL2-7; there is no amino
FT acid change"
FT 1202

Query Match 8.78; Score 178.4; DB 1; Length 1513;
Best Local Similarity 53.08; Pred. No. 1.4e-31;
Matches 405; Conservative 0; Mismatches 356; Indels 3; Gaps 1;

```

OY 629 gaagaaaaactaagctctctggtgaacagcttcgatgcttgagccctgaagcctcaag 688
    |||||
DB 687 GAAGTGTGATGATGCTGTTGCTCAACAGCTTGGAAGGGTTAGAGACACAAAGCTTGTGAA 746
    |||||
OY 689 gccatttaagtaagtaagatgcatcagcgcttgatccctccgactcttgagc 748
    |||||
DB 747 GTTGAGAAATGAGAGGGGATGAGGGCCAGCTTGATGAGACATATGATACCATCTCTTAT 806
    |||||
OY 749 ggttaagatccttcggacaggtcttcggcgagatcttctgagaagaagtcgaatgac 808
    |||||
DB 807 CTGAGGGCCCAATCAAGAGATMAAGGATTAAGGCTGACGGCTGAGCTTGAAAGCCGCTCG 866
    |||||
OY 809 gacgactccctcgaatgcttgagcaagcagcagcctcctcgttggttgcttcgttc 868
    |||||
DB 867 GAGAGGTGATGAGATGAGTGTAGACACTTAAGCTGACGAAGCGGTGTTTCTTCGTTT 926
    |||||
OY 869 ggaagcttcgttaagtaagtaagtcgcaaatggaagatctgcaagaagagcgtctaat 928
    |||||
DB 927 GGTTCCTTTGGGATCTCTTTGAGAACCACTGCTGAGGTGCAAAAGCGTTACAGAA 986
    |||||
OY 929 tctgagagcgctctctgctggtgagtaagtaagcaaggaaggaaglatltaagt 988
    |||||
DB 987 TCCAACTTTAACTTCTGTGGGTGATCAAGAACGTCATATAGCGAAGTTACCAAGAGG 1046
    |||||
OY 989 tgcattgagagagcttggaacagagtgaggaaattgatactctggtctcaatggaatc 1048
    |||||
DB 1047 TTTGTGAGAGCTTACCAAGACAGAGCGCTGCTGTTTGTGTTGATACCACTTACGTT 1106
    |||||
OY 1049 ctgaagcaccctcgtctgagatgcttctgacacacactgagggctggaatcagcctcaag 1108
    |||||
DB 1107 TTAGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1166
    |||||
OY 1109 agtatactcttcggaggttcgagatgcttctcgcagatgcttcgataaagcgagat 1168
    |||||
DB 1167 GATTTGATGATTTGGAGGATTCGATGATGATGATGATGATGATGATGATGATGATGAT 1226
    |||||
OY 1169 gcgaagcctgagagagatgctgagagagcgagctgagagctgagagctcaatgagaggt 1228
    |||||
DB 1227 CCTAGATTGTTGAGAGAGGATTTGAGAGTTGGGTATGAGAGAGAGAGAGAGAGAGAG 1286
    |||||
OY 1229 acgctcgtctgctgagatgagatgagatgagatgagatgagatgagatgagatgagat 1288
    |||||
DB 1287 GGAGTTGTTGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1343
    |||||
OY 1289 aagagtaagaaacttgaagagagctgctgcaagctggaagagatcttgcaagaagaactatg 1348
    |||||
DB 1344 ACTAGTGTGAGATGAGAGAGATTTCAAGAGTTCGAAGAGATTTGCTGTAAGCGCATG 1403
    |||||
OY 1349 gaggagatgagatcttcgataacacactcaagctcttcctga 1392
    |||||
DB 1404 AGTGAAGAGAGAGAGCTCTGATCGAGCATTAATGAGTTGTGGA 1447
    |||||

```

RESULT 7
X02828/c
ID X02828 standard; cDNA: 2062 BP.
AC X02828;
DT 14-MAY-1999 (first entry)
DE M0905287 Seq ID 3.
KW Plant; flavonoid 5-transglycosylation activity; 5T5; variety; colour; ds.
OS Verena hybrida.
FH Key Location/Qualifiers
FT CDS 26..1411
FT /tag= "a
FT /product= "protein with flavonoid 5-transglycosylation
FT activity"
FT
FT M0905287-A1.
FT 04-FEB-1999.
FT 16-JUL-1998; J03199.
FT 25-JUL-1997; JP-200571.
FT (SUNR) SUNTORI LTD.
FT Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
FT

```

PI Yamazaki M;  
DR WPI: 99-142940/12.  
DR P-PSDB: W92949.  
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be  
PT used to transform plants for improvement of plant coloration  
PS Disclosure: Page 56-60; 89pp; Japanese.  
CC This invention describes the isolation of plant proteins which have  
CC flavonoid 5-transglycosylation (5T5) activity. Such proteins can be used  
CC in the production of plant varieties with improved colour, for example in  
CC the production of flower varieties with desired coloration and cut  
CC flowers harvested from them.  
SQ Sequence 2062 BP; 608 A; 354 C; 508 G; 592 T;  

Query Match 7.1%; Score 147; DB 1; Length 2062;  
Best Local Similarity 63.1%; Pred. No. 26-24;  
Matches 330; Conservative 0; Mismatches 125; Indels 68; Gaps 4;  

OY 1433 agcaaaatcgcaaatctcatcctgtaagtgaaatcctttgcttaattgtcga 1492
    |||||
DB 1921 AGCGCAAAATTTGCAAAATTTGCTGAGTAAATTTACTACTTTCGGCTTATTTGCTGGA 1862
    |||||
OY 1493 gttgttttcaaatctt-atctgttaatttcaattgagtggaattagctgattt 1551
    |||||
DB 1861 GTTGTGTTTAAATTTCAATATTAATTTTCAATTTTGTATCAAAAGTTAGTCGATTTT 1802
    |||||
OY 1552 aac-----tgaaaaataataatctgtt 1578
    |||||
DB 1801 GGCTAAAAAAATTCAAAACCTGTTAGAAATTCGGTGAAGAAATTTAGCACTCGTGCCT 1742
    |||||
OY 1579 gagaactcttcaaaaatcatcctgattcttcttctgcttgcgcaaatctccatatca 1638
    |||||
DB 1741 CAGACACTTCATCAAGATCACTGATCACTTATTTATTTATTTGTCAAATTAATACCTCTT 1682
    |||||
OY 1639 atggaaaaaataaatctcaaatctgcacattctgaaccaagaagaagatcaattga 1698
    |||||
DB 1681 CTTGGTTCAAAATTTGAGAGATTTTGAATTTATTTTCCAAATGTAAGAAATTTTGA 1622
    |||||
OY 1699 ccaaaataataaagaatctcaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1758
    |||||
DB 1621 CCAGACATTAAGAAATTCAGATGATTTTGAAGAGAGTCAACAAATGATTTTATAT 1562
    |||||
OY 1759 tttccacgaattctcaagagcttctgaattctttttagccaatctgagcactt 1818
    |||||
DB 1561 TTTTCC-----AGTTAAATGAGACTAAATTT 1535
    |||||
OY 1819 gtacaaaatgaaagctatcatgaaatcttlaaaaaaacacacacacacacacacacacac 1878
    |||||
DB 1534 ACACTCATATGTAATTTCAAGAT-AAATTTTGAAGAAACAACTGCAACAAATTTAAAGC 1476
    |||||
OY 1879 ccgaagaatgaaatctcctgacgaagaattgcaattcgct 1921
    |||||
DB 1475 CAAAAAGTATTCACCTTACGAGATGAGAAATTTGCAATTTTGCCT 1433
    |||||

```

RESULT 8
T66173
ID T66173 standard; cDNA to mRNA; 566 BP.
AC T66173;
DT 15-JUL-1997 (first entry)
DE UDP-glucose:thiolpyridoximate S-glucosyltransferase clone pGL2-7.
KW Glucosinolate; UDP-glucose:thiolpyridoximate S-glucosyltransferase;
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
OS Brassica napus cv. Westar.
FN EP-771878-A1.
PD 07-MAY-1997.
PF 31-OCT-1995; 402425.
PR 31-OCT-1995; EP-402425.
PA (CANADA) NAT RES COUNCIL CANADA.
PA (PLB2) PLANT GENETIC SYSTEMS NV.
PI Grootevassink JMD, Hemmingsen SM, Kolenovsky AD, Peferoen M;
PI Reed DW, Underhill EM, Van Audenhove K;
PI WPI: 97-247418/23.

PT	Plants genetically transformed to interfere with
PT	UDP-glucose:thiohydroximate S-glucosyltransferase gene expression
PR	- useful for production or rapeseed oil with reduced glucosinolate
PR	content
PS	Example 2; Page 18-19; 35pp; English.
CC	Breissia napus UDP-glucosyltransferase
CC	(S-GT) partial cDNA clones: pglu-7 (T66173) and pglu-2.25 (T66174)
CC	were obt'd. by PCR-RACE (see also T66167-72) of B. napus cDNA.
CC	Analysis of the sequences revealed an open reading frame of about
CC	470 bp, a 1043 ' untranslated region and polyA+ tail for both
CC	clones. The amino acid sequence of the encoded protein revealed
CC	part of S-GT peptide 2 (W09827), as expected because this sequence
CC	was used for PCR-cloning, and the complete S-GT peptides 5, 6 (with
CC	one amino acid difference) and 7 (W09830-33). A full-length cDNA
CC	clone (T66166) for B. napus S-GT (W09825) was subsequently obt'd.
SQ	Sequence 566 BP; 160 A; 54 C; 193 G; 159 T;
<hr/>	
Query Match	6.9%; Score 143; DB 1; Length 566;
Best Local Similarity	60.3%; Pred. No. 1e-23;
Matches 255; Conservative	0; Mismatches 165; Indels 3; Gaps 1;
Dy	agaagagatattgataagtgcgatggaggagtggaacgaagtcgggaaatatgtactcttgc 1029
Db	13 AGCGAAATTACAGGAAGGGTTGTGSAAGACTCAAAAGACAGACGCTTGCTGTTCCTTS 72
Oy	gtgttcaccattggaagttccgaagcattcccctcglttgagatggttcgtgacacaacgtcg 1089
Db	73 GTGTAAACCAGCTTGAGGTTTTAGCTATGAGATCTTAAGCTGTTTTTGACTCATCTGCCG 132
Oy	gtggaattcgaccttagagatcatcttcctcgagggttcgcatgtgtgcgtcttcctcgcaagt 1149
Db	133 TTGGAACTCAGACGCTGGAAGGGTGTGAGTTGGCACTTCCGATGGTGGGTGCCGCACAGT 192
Oy	gttcgatccaaggagaagatgycgaagctgatlgtggaagatgttgtgagagacggtgtgagat 1209
Db	193 GAGTGATCAAGATGAATGATGATCTTAAGTTGTGTGASAGAAGSTTTGGAGAGCTTGGATTAAGAC 252
Oy	1210 gagaagcataagagagaggtgagcgtctgtatgtgtaataattagagagatgtatgagga 1269
Db	253 GAAGATGAAAGCTGGGGGAGGAGATTGTGAAGAGCATGAGTGGTGAAGGTTTGAAGAAG 312
Oy	1270 ggttatgtgagtgaggggaaaagagatggaacttgagaagagtgctgycgaatggaaga 1329
Db	313 AGTATATGGA--AGGAGAGAGTAGTGTGGAGATTGAGAAAAGTCTAAGAAAATGGAAAAG 369
Oy	1330 ttgtccaagaaaaagcatctgaggaagaatgatgactctcaatcaacaactcaagtgcttct 1389
Db	370 TTGTGCTGTGAAGCGCATGATGATGAAGAGAGAACGCTGTGATCGGAGCACTTAATGAGTTGT 429
Oy	1390 tga 1392
Db	430 TGA 432
<hr/>	
RESULT	9
ID	T10106
AC	T10106 standard; cDNA to mRNA; 1731 BP.
TI	T10106;
DT	27-OCT-1996 (first entry)
DE	Maize UDP-glucose indol-3-yl:acetylglucosyl transferase sequence.
KM	UDP-glucose-indol-3-yl-acetylglucosyltransferase; enzyme;
KW	transgenic plant; EC-2.4.1.121; crop improvement; corn; cereal;
OS	grass; IAGLU-transferase; ds.
FT	Key mays.
FT	Location/Qualifiers
FT	cds 57..1472
FT	/tag- a
FT	misc_binding 858
FT	/tag- b
PN	/note= "start of UDP binding region"
PD	MO3600291-AI.
	04-JAN-1996.

Query Match	Score 141	DB 1	Length 1731
Best Local Similarity 53.8%			
Matches 315	Conservative 0	Mismatches 265	Indels 6
		Gaps 1	
20-JUN-1995: 007820.			
PR 24-JUN-1994: US-265427.			
PA (RESE) RESEARCH CORP TECHNOLOGIES INC.			
PI Bandurski RS, Szczepkowski K, Szeszen JB;			
DR MPI: 96-068875/07.			
P-PSDB: R99999.			
PR UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein			
PT - used in sense or anti-sense orientation for the genetic control of			
PS plant IAA levels and plant growth			
PS Claim 21: Page 21-23: 41pp; English.			
CC This nucleotide sequence encodes a UDP-glucose-indol-3-			
CC y-lacetylglucosyltransferase (AGLU-transferase) from maize (Zea mays).			
CC Over-expression of this sequence in transgenic plants will result			
CC in at least 1 of the following effects: inhibition of apical			
CC enlargement; inhibition of stem elongation; inhibition of cell			
CC elongation and increased numbers of stems per plant as compared			
CC with a wild-type plant. The ability to control the expression of			
CC this enzyme allows the control of free indoleacetic acid levels in			
CC plants, thereby effecting plant growth rates. Antisense constructs			
CC may be used to inhibit synthesis of this enzyme.			
SO Sequence 1731 BP; 298 A; 528 G; 304 T;			
6.8%			
DB 1			
806 gagcagcattgctcgtgaatggttgagcagcaatccctgcatcttcgtgttgtaacgttcg 865			
DB 849 GAGGATCGGCGACCAACAAAGGCGTAGACACAAAGCCCGACCGCTCCGGGCGCTTCCTCC 908			
866 ttgcgaagctctgttaatacagcagaagtcgcgaatgagagagataagaagagcgctta 925			
DB 909 TTGGGACGCTCCCGCTCCCTGGGCAACGCCCAAGAGAGAGAGCTCGCGCGGCGCTCTCC 968			
926 gattgtggagcgctgttttctgtgtgtgtaagaagtaaaagagaagaagagatttgata 985			
DB 969 GCCCGCCGCAAGCGCTTCCTGTCGTGGCTGAGAGGCGCACGCGACGACCAAGCTCCCGCC 1028			
986 agttgcataagaa-----gttgaacagagtgaggaaatgtaattgtattgtttctaa 1039			
DB 1029 TATCTCCTGGCGCGAGCGAGCGACGACGCGCGCGCGGAGAGTCTGTGCTGCTCCCGCAG 1088			
1040 ttggaagctcttcagcagatccctcgtgtggagattgtctgacacactgcgaggcgtaatcg 1099			
DB 1089 CTGGACCTGCTGGCGGACACCGCGCGGTGGCTTCCTGTCACCACTCCGTTGGAACTCC 1148			
1100 actctagaaggtatattcttcgggggttcgcgaatggtggtttccgcgaagtggttcgtcaa 1159			
DB 1149 ACGGTGAGGCGCTCAGCTTCGCGCTGCTATGTTGTTGAGAGGCGGCTGTGACGACAG 1208			
1160 gggcgaatgcgaagctgtagtgagagatgtgtggaagacgggtgtgtgagatgagagctaat 1219			
DB 1209 CCGACCAACCTCTGGGAACCTCGAGCTCGCTGGGCGCGCGCGCTGCGCGCGCCCGCAT 1268			
1220 gagaagagtgatcgctgtatgtatgtaaatggaagatgatgatgagagtgatgagat 1279			
DB 1269 GCTGCGCGGCGGCTGTTCTTCCTGCGCGGGAATGAGAGCGCTGCTCCGCGCGCTCATGAGAC 1328			
1280 ggggagaaaagagtgagaaacttagagagagtgctctgcgaagtgagagagattggcaaga 1339			
DB 1329 GGGGGCAGAGCGGCTGCTGCTGCACGCAAGAGGAGGGGAGATGAGAGGACAGGCTCCG 1388			
1340 aaagctatgagagaagatgatacttcagattacaacactcaagagcttc 1386			
DB 1389 GCCCGGTGGCACCCGGTGGCAGCTCTGACCGCAACCTGACGAGTT 1435			
RESULT 10			
176174			
ID T66174 standard: cDNA to mRNA: 568 BP.			
AC T66174;			
DT 15-JUL-1997 (first entry)			

QY 1281 gggga-----gaaagagtaggaacttagagagagcgccgcaagtggaaagatt 1331
 Db 1260 GAGTAATGCGAGTGAAGAAACAGAGGCGTTTACAGACGACCAAAAGACTACAAAGATA 1319
 QY 1332 tggcaagaaagctatgaggaagatggatcttc 1365
 Db 1320 TGGCAAGAGAACTATTGAAGAGGAGGATCATC 1353
 RESULT 12
 ID X24873 standard; cDNA: 1626 BP.
 AC X24873;
 DT 21-JUN-1999 (first entry)
 DE Medium-chain UDP-glucose:fatty acid glucosyltransferase cDNA.
 KW UDP-glucose:fatty acid glucosyltransferase; tomato;
 KM 1-O-acyl-a-glucose; acyl donor; esterification; ss.
 OS Lycopersicon pennellii.
 PS Lycopersicon esculentum.
 FT Key Location/Qualifiers
 FT CDS 1..1416
 FT /tag- a
 PN M09090144-A1.
 PF 25-FEB-1999.
 PR 10-AUG-1998; 1B1369.
 PR 30-JUN-1998; US-106464.
 PR 13-AUG-1997; US-055554.
 PA (CORR.) CORNELL RES FOUND INC.
 PI Eanneetta NT, Ghangas GS, Kual J, Steffens JC;
 DR WPI: 99-190155/16.
 P-PSDB: M98009.
 PT New chain length specific uridine diphosphate-glucose:fatty acid
 PT glucosyltransferases - useful for prepping straight or branched,
 PT (un)saturated 1-O-acyl-a-glucoses and 1-O-acyl-a-D-glucoses
 PS Example 3; Page 33-36; 42pp; English.
 CC The present sequence encodes a UDP-glucose:fatty acid
 CC glucosyltransferase that transfers glucose to medium chain length
 CC fatty acids to form 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses
 CC act as acyl donors in the esterification of glucose and in the
 CC further esterification of partially acylated glucose in reactions
 CC catalysed by glucose acyltransferases. They also act as acyl donors
 CC in the esterification of sucrose and in the further esterification
 CC of partially acylated sucrose in reactions catalysed by sucrose
 CC acyltransferases. The source of the present cDNA is leaf trichome
 CC mRNA of an F1 population of a cross between wild tomato
 CC Lycopersicon pennellii (LA 716) and cultivated tomato
 CC Lycopersicon esculentum cv. New Yorker. The clone was isolated
 CC following immunoscreening of a trichome cDNA library. The
 CC invention provides chain length-specific UDP-glucose:fatty acid
 CC glucosyltransferases and a method for their purification. The
 CC enzymes have specificity for transferring glucose to short, medium
 CC or long chain length fatty acids. Methods for preparing
 CC 1-O-acyl-a-D-glucoses, where the acyl group comprises 3-5, 6-13
 CC or 14-22 C atoms, are claimed.
 SO Sequence 1626 BP; 540 A; 256 C; 372 G; 458 T;

OY	984	tatgttgcattgagggaggttgaaacgaagtgggganaaatgtatacttgttc--caat	1040
Db	974	AAGGATTTGGAGGAAGAAGCAAAAAAAGGTTTTAACTATAAGAGTGTGGGCCACCCCAAG	1033
OY	1041	tgsaagtcctgaagcaccctccctcgttcggaatgttcctgatcacactgcgggtgaattcca	1100
Db	1034	TGCATATTCTTGATTAACAAGACTATTGGAGCGTTTGTGTACTCATTTGGTAGTGAACATCCA	1093
OY	1101	ccttagagaatatatcttcggggttcgcgaatggtgcttttcgcagtggttcgatcaag	1160
Db	1094	CACGTGAAGGAATATTCAGCAGGGGTACCAAATTGTACATGTGCCAGTATTTGGCGAACACT	1153
OY	1161	gacacgaatggaagccgcatggaaggatctgctgaggaagcggtgtgagagtgaagctaacg	1220
Db	1154	TTTTTCATGAGAAAGTTGGTGTGACTGAGTAAAGTAAAGAGAGCGTGTGTGTCTTACG	1213
OY	1221	aggaaggtagcgtcgtltagtgatgaatgaagaatgatatgtatgagaggtatagatg	1280
Db	1214	AATGGAAGACMACGTACTGTGAAGGAGTGAACAGCAAGCAATATGCAACGCCATTAAGA	1273
OY	1281	ggsga-----gaaaagtgtaagaaacttagaagaagatgctgcggaagtggaaagatt	1331
Db	1274	GAGTAAATGGCACTGAAGAAACAGAGCGATTACAGACGACGAAAAGAGTACTAAAGCAA	1333
Db	1334	TGGCMAAGAACGCTATTGGAAGGAAGGAGTATC	1367
OY	1332	tgcaagaanaagctatgsgaagaatgtgaccttc	1365
Db	1334	TGGCMAAGAACGCTATTGGAAGGAAGGAGTATC	1367
RESULT	13		
ID	T62123	standard; cDNA to mRNA; 1578 BP.	
AD	T62123;		
DT	10-JUN-1997	(first entry)	
DE	Solanum melongena flavanoid-3-glucosyl transferase cDNA.		
KW	Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;		
KW	piment; production; seed; plant; ds.		
OS	Solanum melongena.		
PH	Key	location/Qualifiers	
FT	cds	13..1314	
FT		/tag=	a
PN	J09056385-A.	/product=	flavanoid-3-glucosyl_transferase
PD	04-MAR-1997.		
PE	25-AUG-1995; 238943.		
PR	25-AUG-1995; JP-238943.		
PA	(KIRI) KIRIN BREWERY KK.		
DR	WPI; 97-20631/19.		
P	-PSDB: W13406.		
PT	Solanum melongena flavanoid-3-glucosyl transferase and DNA encoding		
PT	it - catalyses transfer of glucose to 3-position of flavonoid		
PT	piment in plants		
PS	Claim 2: Page 10: 14pp; Japanese.		
CC	The present sequence encodes Solanum melongena (eggplant)		
CC	flavanoid-3-glucosyl transferase (F3GT), which catalyses the		
CC	transfer of glucose to the 3-position of flavonoid pigments. The		
CC	F3GT gene is useful for the production of the seed of a plant. A		
CC	S. melongena seed was cultured under a red light and a flavanoid		
CC	derived. A cDNA library in which the piment was concentrated and		
CC	prepared. A flavanoid glucosidising enzyme gene was cloned and its		
CC	structure and mRNA expression pattern analysed. The clone was		
CC	expressed in E. coli.		
SO	Sequence 1578 BP; 498 A; 264 C; 310 G; 506 T;		
Query Match	4.7%; Score 96; DB 1; Length 1578;		
Best Local Similarity	48.0%; Pred. No. 6,6e-13;		
Matches 409; Conservative 0; Mismatches 425; Indels 18; Gaps			
OY	804	atgacgaagatgccttcgaatggtgtgagacaagaatcctcgtacttcgggtgttacgttt	863
Db	743	ATGAAATGGATGCATTCATCGCTTGTGACAAACAAAGGAAAAATTCAGTTGTATCTTA	802

CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (sic) petals. The gene is
CC used to transform rose plants to produce blue flowers.
SO Sequence 1597 BP; 483 A; 272 C; 368 G; 474 T;

Query Match 3.2%; Score 66; DB 1; Length 1597;

Best Local Similarity 48.3%; Freq. No. 4.2e-06;

Matches 258; Conservative 0; Mismatches 260; Indels 16; Gaps 2;

```
QY 1028 tgggtctcgaattggaagcctcgaagcctccctcgttgatggttcttgacacacatgc 1087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1056 TGGGCCCCACAGCGCGCGTGTGCTCAGCTGAGGAGGTTGCGTGCATGTG 1115

QY 1088 ggggtgaattgacacctgaagagatatacttcgggggtccgatggtgctttccgcag 1147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1116 GGATGGAAATTCACAGTGAAGCAATGGTGGTGGTCCAGGCAACCTGGCCCTG 1175

QY 1148 tggltcgaatcaaggaagcaatgcgaagctgaaggaagatgltggaagacgggtgaga 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1176 TTGCTGACACAGAGATGATCATTTGATTTGGTGAAGAAATGGGCTTGGCTGGAG 1235

QY 1208 gtaagagcctaagagagaggttagcgtcgtgatagtgaatagaagatgattgag 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1236 GTTAACATGATTTATTAATAAGATTACAGAAATCCTGACGGGAGAAATGTGAGAGCT 1295

QY 1268 gaggtatagatgagagagaaagagtagaaactagaagagagtgctgcaagtgaag 1327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1296 GATGTTATAGAGAGAAATCAAGATTTGATGATCCTGAAATGGAATCAGAAAGAAA 1355

QY 1328 gattgcaagaaaa-----gctatgaggaagatgattcctcaagtaac 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1356 GTGAAGGAAATGMAAAGAAAGACAGGCTGCTGTGAGAAAGAGGATCATCTTGTGCG 1415

QY 1373 aacctcaaggtcttcttgatgaagtgtagtatactaagaagc-taaatgaagtcacca 1431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1416 TCCCTCAAGATTTTATATGATGTATCAAAAGACTCCATGATTTAGCTCTTCCAAA 1475

QY 1432 taagcaaatgcaaatlcaatcgttaagtgaatacttttggcttaatttggctg 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1476 TTGGTTCAATAGATTTTATGGGCAAACTGTTAAGTGTGCTTTGCAATTAGACCAATGC 1535

QY 1492 agttgttttcaaatatctgttaatttacaattgagtgtaatttagtct 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1536 TCTGATTTAATGTAAGTTTTATTTTATTTTATTTATCTGAATTAAGTTT 1589
```

Search completed: August 1, 2000, 18:17:46
Job time: 19987 sec

THIS PAGE BLANK (USPTO)

Query Match 6.8%; Score 141; DB 4; Length 1731;

Best Local Similarity 53.8%; Pred. No. 4,3e-27;
Matches 316; Conservative 0; Mismatches 265; Indels 6; Gaps 1;

```

806 gagcagcattccctcgaatggttgaagcagcaatccctcgtatcttcggtgttaagttcgc 865
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 GAGGATGGGTCCACCAAGTGGCTAGACACCACCCGACCGCTCGGTGGCTTACCTTCC 908
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
866 ttcggaagcttcgttaatacgaagtcgcaaatggaagagatagcaagaggtctgtta 925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
899 TTCGGCACCCCTCGGCTCCCTGGGCAAGCCCAAGAGAGACTGCGCGCGCTCTCTC 968
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
926 gatttggaagccgcttttctgtgtgaagagtaaaaggaaggaagagatattgata 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
969 GCCGCGGCAAGCCGCTTCTGTGGGTGTGAGGCCAGACGACGACACCAAGTCCCGCGC 1028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
986 agtgcacatgagga-----gttgaacgaagtcgggaaaaattgatacttggtgtctaa 1039
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1029 TATCTCTCGGCGGAGGCGACGCGACGCGCGCGCATGATGCTGCTGCTGCTGCGCGCAG 1088
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1040 ttggaagctcgaacatccctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1099
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1089 CTGAGACGTGCTGGGCGACCCGCGCGTGGGCTGCTGTGTCACCCACTGCGGTTGGAACTCC 1148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1100 actctagaagatatacttcggtgttcggaatggtgtgtgtgtgtgtgtgtgtgtgtgt 1159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1149 ACGCTGAGGCGGCTCAGCTTCGCGGCTCTATGTCGATGCGCATGCGCTGTGGACGACAG 1208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1160 gggaagcagcgaagctgcatgagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1209 CCGACCAACGCTCGGAGACGTGACGTGCGCTGGGCGCGGCGGCTGCGCGCGCGCAT 1268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1220 gaggaaggttgcgttcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1269 GCTGGCGGCGGCTGCTTCCTTCGCGGGAAGTGAACGGTGCCTGCGCGCTCTCTGAGAC 1338
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1280 gggaagaaagaaagtaagaaactaagaaagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1329 GGGGCGGAGGCGGCTGCTGTCGACGCAAGGCGGCGGGAATGAGGAGACGCGCTCGC 1388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1340 aaagctatggaagaaagatgattcttcagttaaacactcaaggtctt 1386
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1389 GCCGCGGTGGCACCCGGGTGGACGCTTGACCGCAACCTGAGCAGATT 1435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2

PCT-US95-07820-1
Sequence 1, Application PC/TUS9507820

GENERAL INFORMATION:

APPLICANT: Bandurski, Robert S.
APPLICANT: Szeszen, Jędrzej B.
APPLICANT: Szczygiowski, Krzysztof
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
TITLE OF INVENTION: and Plant Growth.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee & Wanner, P.C.
STREET: 3370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCV/US95/07820
FILING DATE: 19-JUN-1995
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,427

FILING DATE: 24-JUN-1994

ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 11-94B PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080

TELEFAX: 303/499-8089

TELEX: 49617824

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1731 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

FEATURE:

NAME/KEY: CDS

LOCATION: 57..1472

PCT-US95-07820-1

Query Match 6.8%; Score 141; DB 6; Length 1731;

Best Local Similarity 53.8%; Pred. No. 4,3e-27;
Matches 316; Conservative 0; Mismatches 265; Indels 6; Gaps 1;

```

806 gagcagcattccctcgaatggttgaagcagcaatccctcgtatcttcggtgttaagttcgc 865
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 GAGGATGGGTCCACCAAGTGGCTAGACACCACCCGACCGCTCGGTGGCTTACCTTCC 908
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
866 ttcggaagcttcgttaatacgaagtcgcaaatggaagagatagcaagaggtctgtta 925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
909 TTCGGCACCCCTCGGCTCCCTGGGCAAGCCCAAGAGAGACTGCGCGCGCTCTCTC 968
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
926 gatttggaagccgcttttctgtgtgaagagtaaaaggaaggaagagatattgata 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
969 GCCGCGGCAAGCCGCTTCTGTGGGTGTGAGGCGCACGACGACGACCAAGTCCCGCGC 1028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
986 agtgcacatgagga-----gttgaacgaagtcgggaaaaattgatacttggtgtctaa 1039
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1029 TATCTCTCGGCGGAGGCGACGCGACGCGCGCGCATGATGCTGCTGCTGCTGCGCGCAG 1088
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1040 ttggaagctcgaacatccctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1099
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1089 CTGAGACGTGCTGGGCGACCCGCGCGTGGGCTGCTGTGTCACCCACTGCGGTTGGAACTCC 1148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1100 actctagaagatatacttcggtgttcggaatggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1149 ACGCTGAGGCGGCTCAGCTTCGCGGCTCTATGTCGATGCGCATGCGCTGTGGACGACAG 1208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1160 gggaagcagcgaagctgcatgagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1209 CCGACCAACGCTCGGAGACGTGACGTGCGCTGGGCGCGGCGGCTGCGCGCGCGCAT 1268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1220 gaggaaggttgcgttcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1269 GCTGGCGGCGGCTGCTTCCTTCGCGGGAAGTGAACGGTGCCTGCGCGCTCTCTGAGAC 1338
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1280 gggaagaaagaaagtaagaaactaagaaagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1329 GGGGCGGAGGCGGCTGCTGTCGACGCAAGGCGGCGGGAATGAGGAGACGCGCTCGC 1388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1340 aaagctatggaagaaagatgattcttcagttaaacactcaaggtctt 1386
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1389 GCCGCGGTGGCACCCGGGTGGACGCTTGACCGCAACCTGAGCAGATT 1435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 3

US-09-106-464-1

Sequence 1, Application US/09106464

Patent No. 6011145

Page 5

[illegible]

```

1 RESULT 7
2 US-07-991-867B-8
3 : Sequence 8, Application US/07991867B
4 : Patient No. 5476781
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: Moyer, Richard W.
8 : APPLICANT: Hall, Richard L.
9 : APPLICANT: Guidi, Michael E.
10 : TITLE OF INVENTION: No. 5476781e1 Entomopoxvirus Expression System
11 : NUMBER OF SEQUENCES: 66
12 : CORRESPONDENCE ADDRESS:
13 : ADDRESSEE: David R. Saliwanhik
14 : STREET: 2421 N.W. 41st Street, Suite A-1
15 : CITY: Gainesville
16 : STATE: FL
17 : COUNTRY: USA
18 : ZIP: 32606
19 :
20 : COMPUTER READABLE FORM:
21 : MEDIUM TYPE: Floppy disk
22 : COMPUTER: IBM PC compatible
23 : OPERATING SYSTEM: PC-DOS/MS-DOS
24 : SOFTWARE: PatentIn Release #1.0, Version #1.25
25 : CURRENT APPLICATION DATA:
26 : APPLICATION NUMBER: US/07/991,867B
27 : FILING DATE: 12-DEC-1992
28 : CLASSIFICATION: 435
29 : PRIOR APPLICATION DATA:
30 : APPLICATION NUMBER: WO 92/14818
31 : FILING DATE: 12-FEB-1992
32 : PRIOR APPLICATION DATA:
33 : APPLICATION NUMBER: US 07/827,685
34 : FILING DATE: 30-JAN-1992
35 : PRIOR APPLICATION DATA:
36 : APPLICATION NUMBER: US 07/657,584
37 : FILING DATE: 19-FEB-1991
38 : ATTORNEY/AGENT INFORMATION:
39 : NAME: Saliwanhik, David R.
40 : REGISTRATION NUMBER: 31,794
41 : REFERENCE/DOCKET NUMBER: UP114.C3
42 : TELECOMMUNICATION INFORMATION:
43 : TELEPHONE: 904-375-8100
44 : TELEFAX: 904-372-5800
45 : INFORMATION FOR SEQ ID NO: 8:
46 : SEQUENCE CHARACTERISTICS:
47 : LENGTH: 1511 base pairs
48 : TYPE: nucleic acid
49 : STRANDEDNESS: double
50 : TOPOLOGY: unknown
51 : MOLECULE TYPE: DNA (genomic)
52 : ORIGINAL SOURCE:
53 : ORGANISM: Amsacta moorei antemopoxvirus
54 :
55 : FEATURE:
56 : NAME/KEY: CDS
57 : LOCATION: complement (18..218)
58 : FEATURE:
59 : NAME/KEY: CDS
60 : LOCATION: complement (234..782)
61 : FEATURE:
62 : NAME/KEY: CDS

```



```

Db 1324 TAAAGATTAAATTTATAGAAAATTAATTAATTTAAAAAATAGATATATCTTCA 1383
Qy 2039 ttctaaaaaataaaaaaataa 2060
Db 1384 ATGTAAAAAAATAATATACA 1405

RESULT 9
US-08-544-332-8
; Sequence 8, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Gruidi, Michael E.
; TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UPL14.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: 852..1511
US-08-544-332-8

Query Match 2.3%; Score 48.4; DB 4; Length 1511;
Best Local Similarity 42.9%; Pred. No. 0.0027;
Matches 241; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

Qy 1499 ttctaaaattatcttgtaatttcaatgagtgtaattgctgatttaactgga 1558
Db 844 TTATCAAAATGATTTACTTAAATTCGATATATTTTAAATTAATTTAAAAATATATA 903
Qy 1559 aaatatataatcattgttgagaccttcacaaatcacatgattccattatgctc 1618
Db 904 ATTTAAAAAAATTAATTAATTAACAGATTAAGTTTAAATTAATTAATTTAAAAAAT 963
Qy 1619 tggcaaaattccatcatcaatctggaataaataattcaaatcgtccaatgtgacc 1678
Db 964 TAGTTAATTTAGAAAGATTCGATATATATATATGATTAATTAATTTAAATATATTC 1023
Qy 1679 aagaagaagatataatttgacccaaataaagaagattcaagatcctgataagtgct 1738
Db 1024 CAGAAATATTAAAGTTTATATTTTCAAAATTTAAATTTTAAATTTTAAATTTTAA 1083
Qy 1739 ctgagcgaggtctcatattttccaccgaatttcaacgagtttgaatttttta 1798
Db 1084 CAAATTTAAAAATATATACATATTTAGATATCTTATTAACAAAAATAGCAATTAAGTA 1143
Qy 1799 gccaaatcggaactcttgtaaaaataagtaagttatgatagaatttlaaaaaa 1858
Db 1144 ATATTATCTACCACTTCATTAAGATTTTAAATTTGCAATCATGTAATTAATTAAGCT 1203
Qy 1859 aactgagacaataataagccgaagtagtaaatctcagcgaatttgcaattctg 1918
Db 1204 ATATTTTATTAATTAATTTAGTAAATTTTAAATTTAATTAATTAATTAATTTG 1263
Qy 1919 ccctcatltaatttttgggtgtaataaactgggtatttctttaaataa 1978
Db 1264 GTACTTTAATAATGTTTTCCTATTAGTATGTTGAATTAATTAATGCAATCAATCA 1323
Qy 1979 taaagtgaagatgcatgataagcttggtagatataagatgagtgatgtaagat 2038
Db 1324 TAAAGATTTAATAATTTATGAAAAATTAATTAATTTAAAAAATTAATGATATATCTTCA 1383
Qy 2039 ttctaaaaaataaaaaaataa 2060
Db 1384 ATGTAAAAAAATAATATACA 1405

RESULT 10
US-08-975-316-32
; Sequence 32, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSEBERG, Leonard N., HAYUKKALA, Ilka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; THE MODIFICATION OF PLANT LGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316

```

```

: FILING DATE:
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/713,000
: FILING DATE: September 11, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: SLEATH, Janet
: REGISTRATION NUMBER: 37,007
: REFERENCE/DOCKET NUMBER: 11000/1003C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-269-0565
: TELEFAX: 206-269-0563
:
: TELEX:
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 405 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-975-316-32

```

```

Query Match          2.3%; Score 48; DB 4; Length 405;
Best Local Similarity 52.2%; Pred. No. 0.0018;
Matches 167; Conservative 0; Mismatches 135; Indels 18; Gaps 2;

```

```

OY 799 gtcgaatgaacagcagctgcctcgaatggttgagcagaaatcctcgaatcctcgggtgtta 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 ggcgaacacacgagctgcatcttcgctgcgacgaagcccttaactcagctgttga 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 859 cgcttcgttcgaagcttcgaatgaacgaagctgcgaacgaagagatagcaagagg 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 tctgagcttcggagcagcagctgctgtgtaacgagctcggaattttccgaattagctttagg 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 919 gctgttgatctgtgtgagcgcttttctgtgtgtgtgaaga-----gtaaa 963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 tttagccatagccagacgcatctgtgtggttcgaccccggtcagtcagcgcgtc 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 964 cgaaggaagagagatctgaatgaagttgc--atgagagagttgaaacgagttgaggaat 1020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266 ggaactcttagaagatttggcccggtgcttttgagagcattacagagagagggagaaat 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1021 tgtatcttgagtggtcttaattggaagtcctcgaagcattccctcgttgagatgtttcgtgac 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326 tctgaatagggcgctcacaacatgactgctcctcgtcgtgagcgcgttttgac 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1081 acaactgcgggtggaattcga 1100
    ||| ||| ||| ||| |||
DB 386 tcacaaatgagtgaaactca 405

```

```

RESULT 11
US-08-107-755A-1
: Sequence 1, Application US/08107755A
: Patent No. 5721352
: GENERAL INFORMATION:
: APPLICANT: Moyer, Richard W.
: APPLICANT: Hall, Richard L.
: APPLICANT: Gruhl, Michael E.
: TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: Florida
: COUNTRY: U.S.A.
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/107,755A
: FILING DATE: 19-Aug-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/827,658
: FILING DATE: 30-JAN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/657,584
: FILING DATE: 19-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: UFI14.C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (904) 375-8100
: TELEFAX: (904) 372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6768 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Amsacta moorei entomopoxvirus
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: complement (65..1459)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1474..2151
: FEATURE:
: NAME/KEY: CDS
: LOCATION: complement (2239..2475)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2502..2987
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 3080..6091
: FEATURE:
: NAME/KEY: CDS
: LOCATION: complement (6277..6768)
:
: US-08-107-755A-1

```

```

Query Match          2.3%; Score 47; DB 2; Length 6768;
Best Local Similarity 46.2%; Pred. No. 0.012;
Matches 270; Conservative 0; Mismatches 305; Indels 10; Gaps 3;

```

```

OY 1477 ctttaattgttcgagttgttttccaataattcctcgttaatttaccattgagtgtaa 1536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1018 cttctgttttgagatgaaattctgatttgaagaaacatttcacatgaaatgcaagc 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1537 attagctgaatttaactggaagaaacataaattcattgttgagagctcttcacaaat 1596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1078 ctttatatttttttagatgataaatactatttatattttatttttttacttctgtaattt 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1597 catctgattccttattgtctcgtgcaaaattccataatcattggaagaaataaattc 1656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1138 tttttatagctccaaatgaaataaattttcttttaattatttttcaaaatttaattattt 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1657 caaaactgcacaatttgaaccaaagaagataattgaccaaataataaaggat 1716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1198 taattatgatttgatctgaaactgaaactgaaactgaaactgaaactgaaactgaaactg 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1717 tcaagtgacttgatgaagtgctcagcagcagagctcatalatttccacgaattcta 1776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1258 taaatgtaactttacgtctgttcacatgacatgacatgacatgacatgacatgacatgac 1313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1777 acgagtttgaattttttttagcacaatcagacacaaacttggtcacaatgaaagttta 1836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


CORRESPONDENCE ADDRESS:
 ADDRESSEE: Gerard H. Bencen
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/544,332
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/991,867
 FILING DATE: 07-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/107,755
 FILING DATE: 19-AUG-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO 92/14818
 FILING DATE: 12-FEB-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/827,685
 FILING DATE: 30-JAN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/657,584
 FILING DATE: 19-FEB-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Bencen, Gerard H.
 REGISTRATION NUMBER: 35,746
 REFERENCE/DOCKET NUMBER: UF114.C4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 904-375-8100
 TELEFAX: 904-372-5800
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8457 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Amsacta moorei entomopoxvirus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: complement (65..1459)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1474..2151
 FEATURE:
 NAME/KEY: CDS
 LOCATION: complement (2239..2475)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2502..2987
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 3080..6091
 FEATURE:
 NAME/KEY: CDS
 LOCATION: complement (6277..6768)
 US-08-544-332-1

Query Match 2.38; Score 47; DB 4; Length 8457;
 Best Local Similarity 46.28; Prid. No. 0.014;
 Matches 270; Conservative 0; Mismatches 305; Indels 10; Gaps 3;
 1477 cttaatttgctgagttgttttcaaaattatctgttaatttcaattgagtgtaa 1536

Db 1018 CTTCTGTTTGGATTAATTCGATTTTGGAAAAGACCTTCTCATATATTAATGGAATGC 1077
 Qy 1537 attagctgatttaactggaaaacataaaattcattgcttgagactctcatcaaat 1596
 Db 1078 CTTTATATTTTATAGATGTAATAAGATATTAATATATTTATATCTGTGAATAT 1137
 Qy 1597 catcgatttccttattgctctgccaataatctcatcaattggaataaataatt 1656
 Db 1138 TTTTATATGTCAAAATATGAAAAATTTCTTTATATATATTTTCAAAATTAATAT 1197
 Qy 1657 caaatcgctccaatcttgaaacaaagaagataaattgaccaaataataaaggat 1716
 Db 1198 TAATATGATTTGGATCTAAATCAATTAATTAATATATTTCCAGATTTTATAGTA 1257
 Qy 1717 tcaagtgatcttgagaagtgctgagcgagctctatattctccacgaattctca 1776
 Db 1258 TAAATGTTACTTTTACCTCTTCTTTTCATCATCATCATATTTTCTTAATAAGCT 1313
 Qy 1777 acgagcttctgaaatttcttgccaaatcggaactaatttgatcaaaatgaagata 1836
 Db 1314 ---ATATTTGCATTAATATTAATTAATTAATAGCATTTATTAATATACCAT-ATTATCTA 1368
 Qy 1837 tatgctgaattttaaaaaaagaacatcgagaataataaagccgaagtagtaaaata 1896
 Db 1369 TTTTACTTAAATAATACATFACATTAATAATTAATACCAATTTGCGCATTTTAAATTT 1428
 Qy 1897 cctgacgaatttgcaattcgccctcatttaatttcttggtgcttaataatcg 1956
 Db 1429 TATTTGGAATCTCTTAATTTTATTTATTTATTTTATTAATTAATTAATTAATTTTAT 1488
 Qy 1957 gtattactacttaataaataaagtgagatgagctgtagctggtagtatactg 2016
 Db 1489 TTCATACATTTTAAATAATATTTTATTTATTTGCTATTTATAGTATTTATATAT-TAAC 1547
 Qy 2017 agtgcagtgtaagtagcatatttcttaaaaaaataaaaaa 2061
 Db 1548 ATTGCTTTTAAATATAGATTTTAAATTAATTAATTAATTAATTAATTAATTA 1592
 RESULT 14
 US-08-913-374-1/c
 Sequence 1, Application US/08913374
 Patent No. 6057492
 GENERAL INFORMATION:
 APPLICANT: Petrus Theodorus Dehaan
 TITLE OF INVENTION: Virus resistant or tolerant plants
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6057492artls Corporation
 STREET: P.O. Box 12257
 CITY: Research Triangle Park
 STATE: No. 6057492th Carolina
 COUNTRY: US
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/913,374
 FILING DATE: March 22, 1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Hoxie, Thomas
 REGISTRATION NUMBER: 32,993
 REFERENCE/DOCKET NUMBER: 137-1099/PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:

LENGTH: 4821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tosspovirus
US-08-913-374-1

Query Match 2.2%; Score 45.6; DB 5; Length 4821;
Best Local Similarity 49.7%; Pred. No. 0.024;
Matches 145; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

QY 1764 caccgaattcctaagcagatttttgaatttttttagccaaatcgacttaattgtaca 1823
DB 3800 CACTATTATTAAGCTTAATTTCTGTCTATTTTTCATTTTGAATCCAAAACCAACAA 3741
QY 1824 aaatgaagaatataatgaatatttaaaaaaacacacgaacaaataaagccgaa 1883
DB 3740 AATGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGTA 3681
QY 1884 agtagtaaatattcct---gacgaatttcgaatttcgcctcctatttaatttttgg 1940
DB 3680 GGCTGAAAAGCAAACTTTGGTCCGACACTTTTGTGTTTGTGTTTGTATTTTATTT 3621
QY 1941 tctgttaataatcgcgtatttcatttaatttaataaagaatgagatcgatagc 2000
DB 3620 TTTTGTGTTTGTGTTTGTATTTATTTATTTGTTTGTATTTATTTATTTATTTTAA 3561
QY 2001 ttgttgatataatagatgagtgtaagtacgatatcttccaataaaaaa 2052
DB 3560 GATTTTTCTATATATATATATCTCTATATATAGAAATTTGAAATTTTAA 3509

RESULT 15

US-07-867-106-2/c

Sequence 2, Application US/07867106
Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5389526tris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 2.2%; Score 45.4; DB 1; Length 5852;
Best Local Similarity 45.8%; Pred. No. 0.03;
Matches 232; Conservative 0; Mismatches 271; Indels 4; Gaps 2;

QY 1479 ttaatttgctgagtttgcttcaaaattacatgtaatttcatgagtgtaaat 1538
DB 2043 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1984
QY 1539 ctatgctgaatttaactggaataataataatcatgctgagctctcatcaaatca 1598
DB 1983 AGAATAGAAAAAGTGGTAACTACATTAATTTTATAGTTTTCATATTTTAA 1924
QY 1599 tctgatttccttatgcttgcgtcaaatctcatcatgtaattggaataataattca 1658
DB 1923 ATTAATTTTAAATTTGAATTTTAAATTTGATTTTAAATTTGATTTTAAATTTTAA 1864
QY 1659 aaat---cgccaatttgaacaaagaagaatataattgacaaaataataaaga 1715
DB 1863 AAATTTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGTA 1804
QY 1716 ttaaatgctctgtaagaagtgctgagcgaggtctcatatttcccgcaatttct 1775
DB 1803 TTCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1745
QY 1776 aacgaatttgaattttttagccaaatcgacttaattgtacaatgaaagt 1835
DB 1744 AAAAACTAGAAAAATGATGATCAATAGATTTTAAATTTTGTGTTTGTGTTTGT 1685
QY 1836 atatgtagaatttcaaaaacaacacgaacataataaagccgaaagtgaataatt 1895
DB 1684 AAAAAAAGTGCATGACAAAAAAGTGCATGACAAAAAAGTGCATGACAAAAAAG 1625
QY 1896 acctgagaaatttgaatttcgcctcctatttaatttcttgctgcttaataaac 1955
DB 1624 AGGGAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1565
QY 1956 ggtatttacttcaatataaataa 1982
DB 1564 TTTAATTTGAATCATATGATATTACA 1538

Search completed: August 1, 2000, 18:17:13
Job time: 74282 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	1002	42.9	455	2	C71402	probable glucosyl h
2	999.5	42.8	467	2	T02238	glucosyl transferase
3	845	35.2	458	2	C71420	hypothetical prote
4	707	30.3	471	2	A54739	indole-3-acetate h
5	667	28.5	479	2	E71419	probable indole-3
6	639.5	27.4	460	2	T00639	hypothetical protei
7	621.5	26.6	456	2	T00506	indole-3-acetate h
8	613.5	26.3	438	2	T00507	indole-3-acetate h
9	613.5	26.3	484	2	D71419	probable indole-3
10	575.5	22.6	475	2	F71419	probable indole-3
11	570.5	22.4	453	2	T00511	indole-3-acetate h
12	459.5	18.7	453	2	T45603	glucosyltransferase
13	439	18.8	455	2	T00584	indole-3-acetate h
14	435	18.6	438	2	T45602	glucosyltransferase
15	425	18.2	452	2	T12981	hypothetical protei
16	422	18.1	449	2	T45605	glucosyltransferase
17	420.5	18.0	451	2	T45604	glucosyltransferase
18	416.5	17.8	447	2	T12978	hypothetical protei
19	406.5	17.4	440	2	T00583	indole-3-acetate h
20	395.5	15.9	347	2	T06371	probable UDP-glucos
21	388.5	15.6	471	2	S01037	flavonol 3-O-glucos
22	384	15.4	433	2	S5167	glucosyl transferase
23	383.5	15.4	385	2	T12980	hypothetical protei
24	383	15.4	462	2	T01732	UDP-glucose glucos
25	382.5	15.4	471	1	S08325	flavonol 3-O-glucos
26	382.5	15.4	471	1	S01052	flavonol 3-O-glucos
27	379.5	15.2	472	2	T03745	glucosyltransferase
28	377	15.1	452	2	T00981	flavonol 3-O-glucos
29	374.5	15.0	466	2	T07404	probable glucosyl h

30	369.5	15.8	420	2	T08005	flavonol 3-O-glucose
31	368	15.7	448	2	T07786	UDP-glucose
32	364.5	15.6	476	2	T03747	glucosyltransferase
33	361	15.4	472	2	S39507	glucuronosyl transferase
34	356	15.2	452	2	G71416	probable glucosyltransferase
35	355	15.2	478	2	T05423	probable glucosyltransferase
36	354.5	15.2	478	2	T08395	UDP-glucose
37	350	15.0	487	2	S41951	UDP-glucose
38	349	14.9	455	1	X0UBHF	flavonol 3-O-glucose
39	348	14.9	394	2	S41952	UDP-glucose
40	345	14.8	507	2	T45161	glucosyltransferase
41	344.5	14.7	478	2	A71417	hypothetical protein
42	342.5	14.7	481	2	T01850	UDP-glucose
43	339.5	14.5	346	2	S41954	UDP-glucose
44	339	14.5	287	2	S41953	UDP-glucose
45	334.5	14.3	449	2	S41950	UDP-glucose

ALIGNMENTS

RESULT 1
C71402
probable glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: columbiana
C:date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: C71402
R:Byers, M.; Hancock, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; P.; Medler, H.; Medler, E.; Wandtke, R.; Weitenegger, T.; Poll, T.M.; Terryn, N.; G.; Avanaag, T.; Hempel, S.; Kotter, P.; Enliam, K.D.; Rieger, M.; Schaefer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Palmgren
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalmatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113
A:Accession: C71402
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-456 <BEV>
A:Cross-references: GB:J297335; NID:g2244747; PID:e3274440; PID:g2244766
C:Genetics:
A:Map position: 4COP9-4G3845
A:Superfamily: flavonol 03-glucosyltransferase

Query Match	42.98;	Score 1002;	DB 2;	Length 456;
Best Local Similarity	48.18;	Pred. No. 3e-74;		
Matches 218;	Conservative 71;	Mismatches 124;	Indels 40;	Gaps 13;

[illegible]

Db 289 ALTHGVATNRPFMTVIREK---NPEKKKNFLELINGSDB-GLVVGWCSQTAVLAHCA 344
Qy 349 LGCFVTHCGNNAVESISCIPVAVAPQFDDTNAKLEIDAMGTGVRNMEGGVDC 408
Db 345 VGCFTVHCNGNSTLESVPVAPFADPCTTAKLIVETWIRGVKVGEGVDGE 404
Qy 409 EIERCEVMVMDGDKTKLVRENAIKKTLARQA 441
Db 405 EIRRCLEKVMGGEBAEENENKAKAVADA 437

RESULT 2
T02238
glucosyl transferase, jasmonate-induced - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 24-Nov-1999
C:Accession: T02238
R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.
submitted to the EMBL Data Library, January 1997
A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspension
A:Reference number: Z14633
A:Accession: T02238
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-467 <KOJ>
A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901
A:Experimental source: strain BY-2
C:Genetics:
A:Gene: JIGT
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 42.8%; Score 99.5; DB 2; Length 467;
Best Local Similarity 44.4%; Pred. No. 5e-74;
Matches 198; Conservative 83; Mismatches 144; Indels 21; Gaps 7;

Qy 6 VLATPPAGCHINPALOPAKRLKAGTDTFTSYAMRMANASAAAGNPGLDEVAF 65
Db 10 VLIALPFGCHINPSLQFSKLLINLVKYLSSLSAFNRKLNPKI----EGLTFAPF 64
Qy 66 SDGTDGGLKPGGDKR--YSEMKARGSEALRNIL---LNNDVTFVYVSHLFAMAAEVA 120
Db 65 SDGTDGFKSPFDYHLSATKSHSEFIAMLSKAKKGPFRVYITTLIMMASSVA 124
Qy 121 RLSHVPALLMVEPAVLYCIHYHFNGVAD---ETDAGSNETQLRPLSLQBSLPFL 177
Db 125 KRLHLPSTLWIPQATVFDIYYRFTNFANFKYNSDQIETLPLSLSSDPSEFV 184
Qy 178 PATPERRLM--MKKLETLDGEEKAVLVNTFDALPDALTAIDRYELIGPLISAF 235
Db 185 DDVANSNMAVESIKRQELINSENRILVNTFDALDELALRYLKNYMTVIGLIFSSF 244
Qy 236 LDGDPSETSYGDLPEKSEENNCVWLSKRSYVYVSGSVLRPKAOMEIGGL 295
Db 245 LDENDRDNFPAADME--SENNYMEMLDARANKSVIYAFQSVYEISSQMEBISGL 302
Qy 296 ACGRPFLMTREOKNDGEEEBEELSCIGELKMGKIVKSCQLEVLAPALGCVTH 355
Db 303 KCGRPFLMTVRETLNG---EKPEKFLCKDELEKIGRIYRRCSCQMEVLRKSSVGCPLTH 358
Qy 356 CGNNAVESISCIPVAVAPQFDDTNAKLEIDAMGTGVRNMEGGVDCIEHCE 415
Db 359 CGNNTLESLSASGPVACPIAMDQCNKLLIQDWKIGVRVNAKKEGIIKRDFOIC 418
Qy 416 MVMDGDKTKLVRENAIKKTLARQA 441
Db 419 IVMDGAEGEELKRNKMKWDLAKES 444

RESULT 3
C71420
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: C71420
R:Beyan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Weiler, H.; Medler, E.; Mamut, R.; Wetzenecker, T.; Pohl, T.M.; Terry, N.; G
vanagh, T.; Hempel, S.; Kotter, P.; Eutlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Mortfort, A.; Pons, A.; Pujdome
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palmer, K.; Benes, V.; Rechman, S.;
C.; Chaiwattis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; M0ID:98121113
A:Accession: C71420
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-458 <BEV>
A:Cross-references: GB:597339; NID:92244901; PID:e326935; PID:g2244911
C:Genetics:
A:Map position: 4C09-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 36.2%; Score 845; DB 2; Length 458;
Best Local Similarity 41.7%; Pred. No. 2.2e-61;
Matches 194; Conservative 74; Mismatches 129; Indels 68; Gaps 17;

Qy 7 LLATPPAGCHINPALOPAKRLK--AGTDTFTFTSYAN--RMANTASAAAGNPGLDEV 63
Db 15 LFTVPPAGCHINPSLQFSKLLINLVKYLSSLSAFNRKLNPKI----EGLTFAPF 64
Qy 64 AFDGVDGGLKPG--GDGKR-----YSEMKARGSEALRNIL---LNNDVTFVYVSH 111
Db 71 TYSGDHDFGSSAYSXKSDQATGNFMSEMRKGETITLIEDNRKQNPFCVYTTI 130
Qy 112 LFANAAYEARLSHPTALLMVEPAVLYCIHYHFNGYADET---DAGSNETQLRPLSL 167
Db 131 LTVVAELAES-----IFNHFNYEDALISEMANTPSSIRKLPSLP 174
Qy 168 EORSPLPFLPATPERRL--MKKLETLDGEEKAVLVNTFDALPDALTAI--DREYLI 225
Db 175 TVRQIPSTVYSSNYAFILPAFREQIDSLKEINPILNTQLEPPAMSSVDNKRIV 234
Qy 226 GIGPLISAFITDGPSETSYGDLPEKSEENNCVWLSKRSYVYVSGSVLRPKA 285
Db 235 PVGPL-----TLRTDF-----SSRGEYTEMIDTKADSSVLYSGTAVLSK 278
Qy 286 QMEBIGGLACGRPFMTREOKNDGEEEBEELSCIG---ELKMGKIVKSCQ 341
Db 279 QLVETKRLIQRSPFLMTVITDQSYRNKEDQKEB--DCISFRELEIDGIVYVSCDQ 337
Qy 342 EYLAHPALGCVTHCGNNAVESISCIPVAVAPQFDDTNAKLEIDAMGTGVR--RM 399
Db 338 RVLNHRSTGCVTHCGNNTLESLSVGPVAVAPQNDQMNKALKLECKMTGVRVMEK 397
Qy 400 NEGQ--GVDCIEHCEVMVMDGDKTKLVRENAIKKTLARQA 442
Db 398 EEEGVVYVDSEIRCIIEVME--DKAEFRGNATRKDLAEAV 440

RESULT 4
A54739
Indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize
C:Species: Zea mays (maize)
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 29-Oct-1999
C:Accession: A54739
R:Szeresz, J.B.; Szczylowski, K.; Bandurski, R.S.
Science 265, 1699-1701, 1994
A:Title: Iaglu, a gene from Zea mays involved in conjugation of growth hormone indole
A:Reference number: A54739; M0ID:94367368
A:Accession: A54739
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-471 <SZS>


```

OY 65 FSDGVDGGLRGDGRKRYSEMKAEGSEALRNLLNN-----DDVTFFVYSHLFA 114
DB 66 FSDGL-----PKDDPRDHEPL-----TESLRKVGANNNSKIIIEGRKFCIIISVFT--P 112
OY 115 MAALVAVRLSHVPTALLMVEPATVLCIYHFYF---NGYADEIDAGSENEIQLPRLSLQRS 171
DB 113 MWPAVAAMAHNIPCAIILMEACAGFSYRYRYKNTSPFD-LEDPRQKVELGDFLEVRD 171
OY 172 LPTLLPATPERFLMMKKELETLDGEKAKVLYNTFDALPELDAITAI-DKTELIGLPL 230
DB 172 LPTLLPATPERFLMMKKELETLDGEKAKVLYNTFDALPELDAITAI-DKTELIGLPL 230
OY 221 IPSAFLDGEDPSETSYGGDLFEKSEBENNCVEMLNKSPKSSVYVYSGSVLRFPRQMEI 290
DB 230 V-SPEFLGADBDKILDDKSLDMWKADDYCEMLDKQY-----SILKSENOYETI 278
OY 291 CKGLLAGGRPLMMIR---DQKNDGDEEBEELSCIGELKKMKIYKSCQLFVLAHP 347
DB 279 AFAKNGVPEPLMVIIRKKEAEVADYLEDVME-----GQGVVTEWQOQEKILCHM 329
OY 348 ALGCEVTHCGMNSAVESLSCGIPVAVPQMPDQTTNAKLIEDAMCTGVYVAMNMGSGVDG 407
DB 330 AISCFTVHCGMNSTIETVAGVPMVAVPTMFDOPLDARLVDVFGIVGKMKNDV---VDG 386
OY 408 -----CEIERCEVEMVMOGDKTKLYRENAIKMKTLARQAM 442
DB 387 ELKVAEVEKCIDAVTKGTDAADN-RRRAAEIQAOTRSAM 424

```

RESULT 12

```

145603
glucosyltransferase-like protein - Arabidopsis thaliana
N:Alternate names: protein F12A12.180
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C:Accession: T43603
R:Choline, N.; Robert, C.; Brothier, P.; Mancker, P.; Catbolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223008
A:Accession: T45603
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <CHO>
A:Cross-references: EMBL:AL133314
A:Experimental source: cultivar Columbia; BAC clone F12A12
C:Genetics:
A:Map position: 3
A:Introns: 161/1
A>Note: F12A12.180
C:Superfamily: flavonol O3-glucosyltransferase

```

```

Query Match 19.7%; Score 459.5; DB 2: Length 453;
Best Local Similarity 28.3%; Pred. No. 7.6e-30;
Matches 134; Conservative 74; Mismatches 185; Indels 81; Gaps 15;

```

```

OY 3 RRVLLATFPAGHINPALQFAKRLKAGDVTFFTSYVAMRMANTASAAAGNPGLDF 62
DB 7 RRSVYVLPFAOGHISPMQOLATLHLKGFSTIVYOTK-----NYSPSDDTTHDPOF 60
OY 63 VASDGDIDGLKPGDGK-----RYMSEKARSGSEALRNLL-NNDDVTFFVYSHL 112
DB 61 VTIEESL-----ESDPEKNGLPQIPLFKLNKCKKVSFKKCLQOLVLOOSNEISCVIYDF 115
OY 113 FAMAFAVRLSHVPTALLMVEPATVLCIYHFYNGVADEIDA-----GSNEIQLPRLPS 166
DB 116 MYFAFAAKCKCKPLNIIFFTSATACRSYFDKILANNVQAPLKEGTQOGEDELYPEFYP 175
OY 167 LEQSLPTEFLPATPERF---RLMKKELETLDGEKAKVLYNTFDALPELDAITADRYE 223
DB 176 LRYKDFV-----SRFASLESIMEVYRNTVDKRTASSVLIINTASCSLSSFLQOQO 228

```

```

OY 224 L-----IGLPL-----IPSAFLDGEDPSETSYGGDLFEKSEBENNCVEMLNKSPKSSVYV 274
DB 229 LQIPVYPPGPLMHVYASPTSLR-----ENKRSIEMLNOKNNSYIYI 271
OY 275 SFGSVLRFPRQMEIEKGLACGRPLMMIRQKNDGDEEBEELSCIGELKKM--- 331
DB 272 SMGSIALMEINMEIVASGLAASNOHPLMWIRGSIPESEIMSEPE-----EFSKYLD 326
OY 332 -GKIYMSCOLEVLAHPALGCEVTHCGMNSAVESLSCGIPVAVQMPDQTTNAKLIEDA 390
DB 337 RGYLVKMAPQKEVLSHPAVGFWSCGMNSTLESIDQCPMKICRPSGSDQVNNARYECV 386
OY 391 MGTGVYVAMNMGSGVDCEIERCEVE--MYMDGDKTKLYRENAIKMKTLARQAM 442
DB 387 WKIGIOVE---GELDRGVERAVKRLMVDDECEE--MRKRAFSLKQQLASV 433

```

RESULT 13

```

100584
Indole-3-acetate beta-glucosyltransferase homolog T27E13.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00584
R:Rounsley, S.D.; Liu, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome 11 BAC T27E13 genomic sequence.
A:Reference number: 214177
A:Accession: T00584
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-455 <ROU>
A:Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150406
C:Genetics:
A:Experimental source: cultivar Columbia
A:Map position: 2
A:Introns: 166/1
A>Note: T27E13.12
C:Superfamily: flavonol O3-glucosyltransferase

```

```

Query Match 18.8%; Score 439; DB 2: Length 455;
Best Local Similarity 30.1%; Pred. No. 3.6e-28;
Matches 139; Conservative 66; Mismatches 195; Indels 62; Gaps 16;

```

```

OY 4 RRVLLATFPAGHINPALQFAKRLK--AGDVTFFTSYVAMRMANTASAAAGNPGLD 61
DB 12 RHYVAMPYPRGRIHPMNNCKRLVRRYRPLVHTFVYT-EEV-----LGFIGDPPDP 63
OY 62 FVARS---DGYDGLKPGDGRKRYSEMKAEGSEALRNLL-NNDDVTFFVYSHLFA 116
DB 64 RIHSTLPNLIPLSELVAKPFIQIDAIVTRLEEPERKLLDLSNPPPSVYFADTYIMA 123
OY 117 AEVAVRLSHVPTALLMVEPATVLCY-----HFYNGYADEIDAGSENEIQLPRLPS 166
DB 124 VRVGRKRRIIPVASTMTWSATILSFLLSHDLSHGHALPEPEEEV-----VDVYPPGLSP 178
OY 167 LEQSLPTEFLPATPERFLMMKKELETLDGEKAKVLYNTFDALPE--DALTAIDRYE 223
DB 179 TKLRDLP-INDGISDVYFKATLCDELPGAR--SLTTTVAIELEHKAIDAFTSKDIP 235
OY 224 LIGIGPLIPSAFLDGEDPSETSYGGDLFEKSEBENNCVEMLNKSPKSSVYVYSGSVLRF 283
DB 236 VYAIQPLIPF-----EBLSYQND--NKPENYIOWLEQPGSVLYISQGSFSLVS 283
OY 284 KAOMEIEIKGLACGRPLMMIRQKNDGDEEBEELSCIGELKKMKIYKSCOLEV 343
DB 284 EAQMEIEIKGLACGRPLMMIRQKNDGDEEBEELSCIGELKKMKIYKSCOLEV 343
OY 344 LAHPALGCEVTHCGMNSAVESLSCGIPVAVPQMPDQTTNAKLIEDAMCTGVYVAMNMG 403
DB 333 LCHRAVGGFWTHCGFNSTLGIYSVPMIAFPMDIINAKMIVEDIRVGMRIERTKK 392
OY 404 G--VDGCEIERCEVEMVMD--GGDKTKLYRENAIKMKTLARQAM 442

```


Db 393 ELLIGREIKVEYKRMDESEEGKEMRRACDLSEISRGAV 434

RESULT 14

T45602

glucosyltransferase-like protein - Arabidopsis thaliana

N:Alternate names: protein F12A12.170

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000

C:Accession: T45602

R:Choisme, N.; Robert, C.; Brotlier, P.; Mincker, P.; Caltolico, L.; Artiguenave, F.; St

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223008

A:Accession: T45602

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <CHO>

A:Cross-references: EMBL:AL133314

A:Experimental source: cultivar Columbia; BAC clone F12A12

C:Genetics:

A:Map position: 3

A:Introns: 145/1

A:Note: F12A12.170

C:Superfamily: Flavonol O3-glucosyltransferase

Query Match 18.6%; Score 435; DB 2; Length 438;
Best local similarity 28.1%, Pred. No. 7.3e-28;

Matches 124; Conservative 75; Mismatches 194; Indels 48; Gaps 14;

3 RRRVLLATPPAOGHINPALQFAKRLKAGDVTFEFTSVYAMRRMANTASAAAGNPGGLDF 62

8 KRRIVLVPAAGGHTVPMQGLKVLNKGFSITVVEGHF-----NOVSSSSQHPFGPQF 61

63 VAFSDGYDDG--LKPGGDKRYMSMKARGS--EALRNLL--NNDDVTFVYVSHLPAMAA 117

62 VTKESLPSEEEKIGIESMTLNKTSSEAFKDCISQLLQDNCIDACTIIDEVYVPCG 121

118 EVARLSHVTALLWEPATVLCIHYFTNGYADIDAGSNEIQPLRLPSEORSDFTELL 177

122 AAKERSIVSTFSTQSA-----NVSHPDQMDQVVE--NLYPLRYDLPTSGM 169

178 PATPERFRLMKKEKLETLDEGEKAKVIVNFPDALEPDALTAIDREYLIGIPLIPAFID 237

170 GPLDRFELCR---EVANKRTASAVIINTVSCLESSSLMLQKGVISYPLGLPHMD 225

238 GEDPSETSGDLPEKSEENNCVEMLNKPKSSVYVYSGSVLRFPAQMEIEIGKILAC 297

226 SSPSS-----ELE--EDRSCIEWLNKQKPSVIYISITGLQMETKEVLEMSWGLCNS 276

298 GRPPLMIREQK--NDDEEEEBEELSCIGELKKMKIYVSWCSOLEVLNHPALGCVTH 355

277 NQDFLWVIRAGSIILGTNGIESLPEDVNMVSE---RGYIVKRAQIEVLGHPAVGFWSH 333

356 CGNNSAVESLSCGIVVAVPQMFQDTNAKLIEDAKGTGVAVRNMEGGVDCGEIERCV 415

334 CGNNSTLISIGVGPICKPFHGEOKLNAMYIESVWRIGFOVE---GKVDGEVERAVK 389

416 -WVMDGDKTKLVRENAIKWK 435

390 RLIVD--DEGAGMRERALVVK 408

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

A:Accession: T12981

A:Molecule type: DNA

A:Residues: 1-452 <CHO>

A:Cross-references: EMBL:AL096859; GSPDB:GNC00061; ATSP:T6H20.280

A:Experimental source: cultivar Columbia; BAC clone T6H20

C:Genetics:

A:Gene: ATSP:T6H20.280

A:Map position: 3

A:Introns: 161/1

C:Superfamily: Flavonol O3-glucosyltransferase

Query Match 18.2%; Score 425; DB 2; Length 452;
Best local similarity 28.3%, Pred. No. 5.1e-27;

Matches 128; Conservative 82; Mismatches 190; Indels 52; Gaps 16;

3 RRRVLLATPPAOGHINPALQFAKRLKAGDVTFEFTSVYAMRRMANTASAAAGNPGGLDF 62

7 KRRIVLVPAAGGHTVPMQGLKALQSGPLIT-----VAQROFNQIGSSLOHFGPQF 60

63 VAFSDGYDDGDKPGGDKR-----YSEM---KARGSEALRNL--LLNDDVTFVYVSHL 112

61 VTIPESL-----PQSESKLQPAEYTLNLTSEAFRECSLSMQGNDICTIYDKL 115

113 FAMAEEVARSHTVPTALLWEPATVLCIHYFNGYADE--IDAGSNEIQ--LPLPLSL 167

116 MYCEAAAKEFKIPSVIFSTSSATIQVCYVLSLSAEKFLIDMKDEKODKYLEGLHPL 175

168 EORSLEPLF--LLPATPERFLMKKEKLETLDEGEKAKVIVNFPDALEPDALTAIDREYLIG 226

176 RYKDLPLSGEPLER-----LLEMKREYVNNKRTASAVIINTVSCLESSSLMLQKGVISYPLGLPHI 230

227 IGPLIPSAFLDGEDPSETSYGDLPEKSEENNCVEMLNKPKSSVYVYSGSVLRFPAQ 286

231 VYPLGLPHI-----TASSPGPSLQ--EDMSCIEWLNKQKPSVIYISITGLRAHMETKE 282

287 MEETIGGLACGRPLMIREQKNDG--EBEDEEBELSCIGELKKMKIYVSWCSOLEVLNHPALGCVTH 344

283 MLEKAMGLNLSNPLWVIRPQSVAGFEWIELPEVYKVAE---RGYIAKRAQIEVLGHPAVGFWSH 339

345 AHPALGCVTHCGNNSAVESLSCGIVVAVPQMFQDTNAKLIEDAKGTGVAVRNMEGGVDCGEIERCV 404

340 GHPAYGCFWGHCGNNTLISIEGVPMICRPLQGGOKLNAMYIESVWRIGFOVE---GE 395

405 VDCGEIERCV--WVMDGDKTKLVRENAIKWK 435

396 VERBGEVRAVKRLIID--EAGAAMERALDLK 425

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Search completed: August 1, 2000, 21:27:45
Job time: 11213 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:31:33 ; Search time 61.65 Seconds
(without alignments)
498.216 Million cell updates/sec

Title: US-09-147-955-4

Perfect score: 2337

Sequence: 1 MVRRLVLLATFPAQGHINPA.....TKLVRENAIKWTLRQAMG 443

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL12:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.todent:*
- 12: sp.virus:*
- 13: sp.unclassified:*
- 14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2337	100.0	443	10 Q9ZR26	Q9ZR26 perilla fru
2	2223	95.1	460	10 Q9ZR27	Q9ZR27 perilla fru
3	1583.5	67.8	461	10 Q9ZR25	Q9ZR25 verbena x h
4	1002	42.9	456	10 Q23270	Q23270 arabidopsis
5	999.5	42.8	467	10 P93709	P93709 nicotiana t
6	940	40.2	519	10 Q9ZVY2	Q9ZVY2 arabidopsis
7	923	39.5	474	10 Q04930	Q04930 arabidopsis
8	914.5	39.1	455	10 Q9ZVY5	Q9ZVY5 arabidopsis
9	845	36.2	458	10 Q23406	Q23406 arabidopsis
10	667	28.5	479	10 Q23401	Q23401 arabidopsis
11	657.5	28.1	449	10 Q22822	Q22822 arabidopsis
12	644.5	27.6	449	10 Q22820	Q22820 arabidopsis
13	639.5	27.4	460	10 Q48676	Q48676 arabidopsis
14	621.5	26.6	456	10 Q22182	Q22182 arabidopsis
15	613.5	26.3	438	10 Q22183	Q22183 arabidopsis
16	613.5	26.3	484	10 Q23400	Q23400 arabidopsis
17	575.5	24.6	475	10 Q23402	Q23402 arabidopsis
18	570.5	24.4	453	10 Q22186	Q22186 arabidopsis
19	551	23.6	481	10 Q9ZWJ3	Q9ZWJ3 arabidopsis

20	458	20.0	482	10 Q9ZUV0	Q9ZUV0 arabidopsis
21	439	18.8	455	10 Q64733	Q64733 arabidopsis
22	426.5	18.2	477	10 Q9ZWQ5	Q9ZWQ5 vigna mungo
23	406.5	17.4	440	10 Q64732	Q64732 arabidopsis
24	399.5	17.1	455	10 Q9ZWS2	Q9ZWS2 vigna mungo
25	396.5	17.0	491	10 Q9ZQ99	Q9ZQ99 arabidopsis
26	395.5	16.9	347	10 Q49042	Q49042 pisum sativ
27	395.5	16.9	496	10 Q9ZQ97	Q9ZQ97 arabidopsis
28	388	16.6	496	10 Q9ZQ96	Q9ZQ96 arabidopsis
29	383	16.4	454	10 Q9XF16	Q9XF16 forsythia x
30	383	16.4	462	10 Q04622	Q04622 arabidopsis
31	383	16.4	495	10 Q9ZQ95	Q9ZQ95 arabidopsis
32	382	16.3	496	10 Q9ZQ98	Q9ZQ98 arabidopsis
33	380	16.3	484	10 Q9ZQ94	Q9ZQ94 arabidopsis
34	379.5	16.2	476	10 P93364	P93364 nicotiana t
35	377	16.1	452	10 Q48715	Q48715 arabidopsis
36	374.5	16.0	466	10 Q43526	Q43526 lycopersico
37	374.5	16.0	474	10 Q82382	Q82382 arabidopsis
38	374	16.0	447	10 Q04114	Q04114 perilla fru
39	373.5	16.0	495	10 Q9ZQ94	Q9ZQ94 arabidopsis
40	369.5	15.8	420	10 Q22603	Q22603 ipomoea pur
41	368.5	15.8	280	10 Q9ZWQ4	Q9ZWQ4 vigna mungo
42	368	15.7	488	10 P93789	P93789 solanum tub
43	367.5	15.7	481	10 Q82381	Q82381 arabidopsis
44	364.5	15.6	476	10 P93365	P93365 nicotiana t
45	359	15.4	381	10 Q9ZWQ3	Q9ZWQ3 vigna mungo

ALIGNMENTS

RESULT 1

Q9ZR26 ID Q9ZR26 PRELIMINARY: PRT: 443 AA.

AC Q9ZR26; DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE HOMOLOGUE.

GN PF3R6

OS Perilla frutescens

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.

OC [1]

RN SEQUENCE FROM N.A.

RC STRAIN=CV. SHIKUN; TISSUE=LEAF;

RX MEDLINE: 90167509;

RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,

RA KUSUMI T., SAITO K.;

RT "Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin.";

RL J. Biol. Chem. 274:7405-7411(1999).

DR EMBL: AB013597; BAA36422.1; .

DR PROSITE: PS00375; UDPGT; 1.

SW TRANSFERASE.

QY SEQUENCE 443 AA; 49110 MW; D87B5486 CRC32;

Query Match 100.0%; Score 2337; DB 10; Length 443;
Best Local Similarity 100.0%; Pred. No. 2.1e-185;
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRRLVLLATFPAQGHINPALQFAKRLKAGTDVFTTSYVAVRRMANTASAAAGNPGL 60

Db 1 MVRRLVLLATFPAQGHINPALQFAKRLKAGTDVFTTSYVAVRRMANTASAAAGNPGL 60

QY 61 DFVAFSDGVDGGLKPGDGKRYMSEMKARGSEALNLLNDDVTFVYVSHLFAWAEVA 120

Db 61 DFVAFSDGVDGGLKPGDGKRYMSEMKARGSEALNLLNDDVTFVYVSHLFAWAEVA 120

QY 121 RLSHVPTALLWVEPATVLCIYHFPNGYADEIDAGSNEIQPLRPLSEQRSLPTFLLPAT 180

```

Db 121 RUSHVPTALLWVEATVLCIYHFYNGYADEIDAGSNEIQRLPSLBOORSUPTFLPAT 180
QY 181 PERFLMMKEKLETLDGEEKAKVLNFTDALEPDALTAIDRYELIGIGLIPSAFLDGED 240
Db 181 PERFLMMKEKLETLDGEEKAKVLNFTDALEPDALTAIDRYELIGIGLIPSAFLDGED 240
QY 241 PSETSYGDLPEKSEENNCVWLNPKKSSVYVSGVSLRPFPPKQMEIEIGKGLACGRP 300
Db 241 PSETSYGDLPEKSEENNCVWLNPKKSSVYVSGVSLRPFPPKQMEIEIGKGLACGRP 300
QY 301 FLWMIREQKNDGDEEBEELSICIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNS 360
Db 301 FLWMIREQKNDGDEEBEELSICIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNS 360
QY 361 AVESLSCGIPVAVPQWFDQTTNAKLIEDAWGTGVRVNRNMGVGGVDCGEIERCEVMVMDG 420
Db 361 AVESLSCGIPVAVPQWFDQTTNAKLIEDAWGTGVRVNRNMGVGGVDCGEIERCEVMVMDG 420
QY 421 GDKTKLVRENAIKWKTILARQAMG 443
Db 421 GDKTKLVRENAIKWKTILARQAMG 443

RESULT 2
Q9ZR27 ID Q9ZR27 PRELIMINARY; PRT; 460 AA.
AC Q9ZR27;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN PF3R4.
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiales; Perilla.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SHIKUN; TISSUE=LEAF;
RX MEDLINE: 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL: AB013596; BAA36421.1; -.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 460 AA; 50974 MW; FE7CCF22 CRC32;

Query Match 95.1%; Score 2223; DB 10; Length 460;
Best Local Similarity 95.7%; Pred. No. 5.9e-176;
Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;

QY 1 WYRRVLLATPPAOGHINPALQFAKRLKLLKAGTDVTFSTVYAWRRMANTASAAAGNPGL 60
Db 1 WYRRVLLATPPAOGHINPALQFAKRLKLLKAGTDVTFSTVYAWRRMANTASAAAGNPGL 60
QY 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYVSHLFAAEVA 120
Db 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYVSHLFAAEVA 120
QY 121 RLSHVPTALLWVEATVLCIYHFYNGYADEIDAGSNEIQRLPSLBOORSUPTFLPAT 180
Db 121 RESQVPSALLWVEATVLCIYHFYNGYADEIDAGSNEIQRLPSLBOORSUPTFLPAT 180
QY 181 PERFLMMKEKLETLDGEEKAKVLNFTDALEPDALTAIDRYELIGIGLIPSAFLDGED 240
Db 181 PERFLMMKEKLETLDGEEKAKVLNFTDALEPDALTAIDRYELIGIGLIPSAFLDGED 240

```

```

QY 241 PSETSYGDLPEKSEENNCVWLNPKKSSVYVSGVSLRPFPPKQMEIEIGKGLACGRP 300
Db 241 PSETSYGDLPEKSEENNCVWLNPKKSSVYVSGVSLRPFPPKQMEIEIGKGLACGRP 300
QY 301 FLWMIREQKNDGDEEBEELSICIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNS 360
Db 301 FLWMIREQKNDG--EBEEELSICIGELKMGKIVSWCSQLEVLALHPALGCFVTHCGWNS 358
QY 361 AVESLSCGIPVAVPQWFDQTTNAKLIEDAWGTGVRVNRNMGVGGVDCGEIERCEVMVMDG 420
Db 359 AVESLSCGIPVAVPQWFDQTTNAKLIEDAWGTGVRVNRNMGVGGVDCGEIERCEVMVMDG 418
QY 421 GDKTKLVRENAIKWKTILARQAMG 443
Db 419 GEKSLVRENAIKWKTILAREAMG 441

RESULT 3
Q9ZR25 ID Q9ZR25 PRELIMINARY; PRT; 461 AA.
AC Q9ZR25;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN HGT8.
OS Verbena x hybrida.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;
OC Verbena.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PETAL;
RX MEDLINE: 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL: AB013598; BAA36423.1; -.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 461 AA; 51347 MW; 6DCBC3B6 CRC32;

Query Match 67.8%; Score 1583.5; DB 10; Length 461;
Best Local Similarity 68.7%; Pred. No. 4.8e-123;
Matches 310; Conservative 57; Mismatches 65; Indels 19; Gaps 7;

QY 1 WYRRVLLATPPAOGHINPALQFAKRLKLLKAGTDVTFSTVYAWRRMANTASAAAGNPGL 60
Db 1 MSRAHVLLATPPAOGHINPALQFAKRLANADIQVTFSTVYAWRRMSRT---AAGSGLI 57
QY 61 DFVAFSDGYDGLKPGDGKRYMSEMKARGSEALRN-LLNNDV----VTFVYVSHLFA 114
Db 58 NFVFSDDGYDGLQPGDDGKYNSEMSKRGKALSOTLANVNDQKSSITFVYVSHLFA 117
QY 115 WAAEVARLSHVPTALLWVEATVLCIYHFYNGYADEIDAGSNEIQRLPSLBOORSUPL 173
Db 118 WAAKVAREFHLRSALLWTEPATVLDIFVYFNGYSDEIDAGSDAIHLPGCLVLAQRDL 177
QY 174 TELLPATPERFLMMKEKLETLDGEEKAKVLNFTDALEPDALTAIDRYELIGIGLIPS 233
Db 178 SELLSTHERPSLMKEKLETLEGEKPKVLVNSFPALEPDALKAKDKYEMTAIGLIPS 237
QY 234 AFLDGEDSETSYGDLPEK--SEENNCVWLNPKKSSVYVSGVSLRPFPPKQMEIEIGK 292
Db 238 AFLDGDKPSDRKSGDLPEKSGDDDDCLEHLSLNPRSSVYVSGVSVNNTTKSOMBEIAR 297
QY 293 GLLAGCRPFLMWIREQKNDGDEEBEELSICIGELKMGKIVSWCSQLEVLALHPALGCF 352

```

```

Db 298 GLLDGCRPLATVR-----VNGEEVLSCBDELKRVCKIVSWCSQLEVLTHPSLQCF 350
QY 353 VTHCGNNSAVESLSCGIPVAVPQWFOQTNNAKLIEDAMGTGVRVRNREGG-VDCETE 411
Db 351 VTHCGNNSLIESIFGVPWAFQWFOQTNNAKLIEDAMGTGVRVRNREGG-VDCETE 410
QY 412 RCVEVMYDGGDKTKLVRENAIKWTKLARQAM 442
Db 411 RCIEEVMDCGCKSRKLDRESAGKWKDLAKAM 441

RESULT 4
ID O23270 PRELIMINARY: PRT: 456 AA.
AC O23270;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)
DE GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERGAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,
RA WEDLER E., WAMBURTT R., WEITZENEGGER T., POHL T.M., TERRY N.,
RA GLENN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVET M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOURELATOU E., MILIONI D., HATZPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HILBERT K., DUESTERHOFF A., MOORES T., JONES J.D.G., ENOVA T.,
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;
RT *Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.;
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z97335; CAB0189.1; -.
DR PROSITE: PS00375; UPGGT; 1.
DR PFAM: PF00201; UPGGT; 1.
KW transferase.
SQ SEQUENCE 456 AA; 50702 MW; EAEDD380 CRC32;

Query Match 42.9%; Score 1002; DB 10; Length 456;
Best Local Similarity 48.1%; Pred. No. 6,le-75;
Matches 218; Conservative 71; Mismatches 124; Indels 40; Gaps 13;

QY 3 RRVLLATFPAGCHINPALQFAKRLKAGTQVTFSTSVYARRMANTASAAAGNPP---G 59
Db 11 RPHYLLVTFPAGCHINPALQANLRIHHGATVYTSVSAHRRM-----GEPSTKG 62
QY 60 LDFVPSDGYDGLKPGDGKR-YMSEMKARGSEALRNLLNND-----VFVYVSHL 112
Db 63 LSFAMTFDGDGLKSFEDOKYNSLRCGNSALRDIKANLDATETEPTGVYISVL 122
QY 113 FAWAEVARKLSHVPTALWVEPATVLCIYHYFN-GVADEIDAGSNIQPLRPSLEQRS 171
Db 123 VPMVSTVAREFHLPITLLWIEPATVLDIYYFYNTSYKHLFDV--EPILKPLKITGD 180
QY 172 LPTFLP--ATPERFLMKKELETLDGEEKAKVLVNTFDALPDALTADRYELIGIP 229
Db 181 LPSFLQPSALFSAL-VTLREHIEALETESNPKILVNTFSALEDALTSVEKLMPIGP 239

```

```

QY 230 LIPSALFDGEDPSETSYGGDLFEKSEENNCVWLNPKPKSVVYVSEGS-VLRFPKQAE 288
Db 240 LVSSS--EGKT-----DLFKSSDE--DYTKWLDOSKLSRSVIYISLGHADDLPEKHE 288
QY 289 EIGKGLLACGRPLAWIREOKNDGCEEEELSCIGELKMKGIYVSWCSQLEVLHAP 348
Db 289 ALTHGVLATNREFLWIVREK---NPEEAKKRFLELRGSDR-GLVYVNCSTAYLAHCA 344
QY 349 LGCFTVTHCGNNSAVESLSCGIPVAVPQWFOQTNNAKLIEDAMGTGVRVRNREGGVDGC 408
Db 345 VGCFTVTHCGNNSLIESLSCGIPVAVPQWFOQTNNAKLIEDAMGTGVRVRNREGGVDGC 404
QY 409 ETIERCEVMYDGGDKTKLVRENAIKWTKLARQ 441
Db 405 EIRRCLEKLVMSGGERAEEMRENAEKWAKAMAYDA 437

RESULT 5
ID P93709 PRELIMINARY: PRT: 467 AA.
AC P93709;
DT 01-MAY-1997 (TREMREL. 03, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN JIGT.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BY-2; TISSUE-SUSPENSION CULTURE;
RA KOJIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB000623; BAA19155.1; -.
DR MENDEL: 9421; Nicta:1525;9421.
DR PROSITE: PS00375; UPGGT; 1.
DR PFAM: PF00201; UPGGT; 1.
KW transferase.
SQ SEQUENCE 467 AA; 52989 MW; 3726EB31 CRC32;

```

```

Query Match 42.8%; Score 999.5; DB 10; Length 467;
Best Local Similarity 44.4%; Pred. No. 1e-74;
Matches 198; Conservative 83; Mismatches 144; Indels 21; Gaps 7;

QY 6 VLLATFPAGCHINPALQFAKRLKAGTQVTFSTSVYARRMANTASAAAGNPPGLDPAF 65
Db 10 VLLATFPAGCHINPALQFAKRLKAGTQVTFSTSVYARRMANTASAAAGNPPGLDPAF 64
QY 66 SDGYDGLKPGDGKR-YMSEMKARGSEALRNLL-----LNNDVTFVYVSHLFAWAEVA 120
Db 65 SDGYDGNFKGSPDDYHLFNSAKSHGSEFIANLIKSAKAKGYPFTRVIYITLMDAGSVA 124
QY 121 RLSHVPTALWVEPATVLCIYHYFNAGYAD---EIDAGSNIQPLRPSLEQRSPLTFL 177
Db 125 KLUHIFSTFWIQPATVFDIYYFYNTFNTFNTFNTFNTFNTFNTFNTFNTFNTFNTFNTF 184
QY 178 PATPERFLM--MKEKLETLDEGEAKVLVNTFDALPDALTADRYELIGIPLSAF 235
Db 185 DDYKSNDAWVESIKROETILNSENPRILVNTFDALNLRVLKNVTMVGIGLIPSSF 244
QY 236 LDGEDPSETSYGGDLFEKSEENNCVWLNPKPKSVVYVSEGS-VLRFPKQAEIKGL 295
Db 245 LDEKDRKDNFFAADMTE--SENNYMEWLDARANKSVIYAFGSYAEISSOMMEEISG 302
QY 296 ACGRPFLMWIREOKNDGCEEEELSCIGELKMKGIYVSWCSQLEVLHAPALGCFVTH 355
Db 303 KCGRPFLWVIRETLN-----EKPEEKLCCKDEKIGRIVNCSQMEVLKSHSSVGCPLTH 358

```

```

QY 356 CGWNSAVESLSCGIPVAVPWFQDQTTNAKLIEDAWGTGVRVNMGGVDCGCEIERCYE 415
DB 359 CWNSTLSEASGVPVIGVAVPWINDQICNAKLQDVKWKGVRNANKEJTIKRDEPKQRCIE 418

QY 416 MVMGGDKTKLVRENAIKWTKLARQA 441
DB 419 IVMGDAEGBELRKNQAKWDLAKES 444

RESULT 6
Q9ZVY2 PRELIMINARY; PRT; 519 AA.
AC Q9ZVY2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE T25N20.20
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BURHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OUI O., SHEN Y.K., TORIUMI M.,
RA VYSOTSKAYA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT *Genomic sequence for Arabidopsis thaliana BAC T25N20.*;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80599.1;
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 519 AA; 58584 MW; C982007F CRC32;

```

```

Query Match 40.2%; Score 940; DB 10; Length 519;
Best Local Similarity 44.3%; Pred. No. 9.9e-70;
Matches 199; Conservative 83; Mismatches 133; Indels 32; Gaps 11;

QY 7 LLAFTPAAGHINPALQFAKRLK--AGTDVTFSTVYAMRMANTASAAAGNPPGLDFVAF 65
DB 57 LLVTFPAAGHYNPSLRARRLIKTGARTFTVCVSFH--NSMIANHNKVENLSFTLF 113

QY 66 SDGYDD-GLKCGGCKRYMSEMKARGSEALRNLL---LNDD-VTFVYVSHLFAHAEVA 120
DB 114 SDGFDGCGISTYEDQRKRSVNLKVGNGDKALSDPIEATKNGDSPVTCLYTILLNAPKVA 173

QY 121 RLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNEIQLPRLPSLQPSLPTLPAT 180
DB 174 RFQQLPSALLWQALVFNIIYTHFMG-----NKSVEFLNLSLEIRDLPSLTPSN 226

QY 181 PRFRFL-MMKEKLETLGEEKAKVLTNTDPALEPDALTAIDRYELIGIGLIPSLFDGE 239
DB 227 TNKGAYDAFQEMNEFLIKETPKILINTDPSLEPEALTAPPNIDMVAVGGLPTIEFSGS 286

QY 240 DPSETS YGDLFEKSENNCVMENLNSKPSVVYVSGSVLFRPPKAQMBEIGKGLLAGCR 299
DB 287 TNKSV-----KQSSSYTLWDSKTESVYVSGFTWVLSKKQIEELARALIEGR 338

QY 300 PFLMIRQKN-----DQGESEEEELSCI-GELKMGKIVSWCSOLEVLAHPALGCFVT 354
DB 339 PFLAWITDKSNRETKTEGEETEIEKTAGFRHELEVGMIVSWCSQIEVLSHRAVGCFTV 398

QY 355 HCGWNSAVESLSCGIPVAVPWFQDQTTNAKLIEDAWGTGVRVNMGGVDCGCEIERCY 414
DB 399 HCGWNSSTLESVLGVPVAVPWFMSDQPTNAKLEESKMTGVRVRENKDLGVERGEIRCL 458

QY 415 ENVMGGDKTKLVRENAIKWTKLARQA 441
DB 459 EAVME--EKSVELRENKAKWKLAMEA 483

RESULT 7

```

```

004930 PRELIMINARY; PRT; 474 AA.
AC 004930;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
DE (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE).
GN IAGLU
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA GILAD A., SCOLNIK P.A., BAR-ZVI D.;
RL Plant Physiol. 113:1004-1004(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA THORNBURG R.W., GRAHAM R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOLE-3-ACETATE -> UDP +
CC INDOLE-3-ACETYL-BETA-1-D-GLUCOSE.
CC EMBL; U81293; AAB58497.1;
DR MENDEL; 16583; Arath; 2543; 16583.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW transferase; Glycosyltransferase.
SQ SEQUENCE 474 AA; 53740 MW; 35521D70 CRC32;

Query Match 39.5%; Score 923; DB 10; Length 474;
Best Local Similarity 43.9%; Pred. No. 2.2e-68;
Matches 204; Conservative 77; Mismatches 133; Indels 52; Gaps 16;

QY 7 LLAFTPAAGHINPALQFAKRLK--AGTDVTFSTVYAM--RMANTASAAAGNPPGLDFV 63
DB 15 LFTVTPAAGHINPSLEAKLAGTISGARVTFPAASISAYNRRMFSTENV---PETLIFA 70

QY 64 AFDGVDGDLKPG--GDGK-----YMSMKARGSEALRNLL---LNDDVTFVYVSH 111
DB 71 TYSOGHDDGFTSSAYSKSDODATGFMSEMRGRRETLEIEDNRKONRPFCTCVITI 130

QY 112 LFANAAEVARLSHVPTALLWVEPATVLCIYHFYFNGYADEI---DAGSNEIQLPRLPSL 167
DB 131 LLTWVAELAREFLPSALLWQPVTFESIFYHYFNGYDAISEMANTFSSSIKLPSLPL 190

QY 168 EORSPLTLLPATPERFL-MMKEKLETLGEEKAKVLTNTDPALEPDALTAI-DRVELI 225
DB 191 TVRDIPSIVSSNYAFLLPAFREQIDSLAKEINPKILINTFOELEPEAMSVDPNFKIV 250

QY 226 GIGLIPSLAFDGEDPSETS YGDLFEKSENNCVMENLNSKPSVVYVSGSVLFRPPKA 285
DB 251 PVGPLL-----TLRTDF-----SSRGEY IEWLDTKADSSVLYVSGTLAVLSKK 294

QY 286 QMEETGKLLACGRPFLWMTREQNDGDEEEELSCIGELKK---MGIKIVSWCSOL 341
DB 295 QVLEUCKALISRRPFLWVITDKSYRNKEDEKEEB-DCISSEKSFDEIGWVSNCDQF 353

QY 342 EYLAHPALGCFVTTCGWN SAVESLSCGIPVAVPWFQDQTTNAKLIEDAWGTGVRV--RM 399
DB 354 RVLNHRSGCFVTHCGWNSTLESVLGVPVAVPWFQNDQNTNAKLIEDCWTGVRVMEKK 413

QY 400 NMSG--GVDGCEIERCYEMVMDGDKTKLVRENAIKWTKLARQA 442
DB 414 EEEGVVVVDSEIRRCIEEWE--DKAEFRGNATRWKDLAAEAV 456

RESULT 8
Q9ZVY5 PRELIMINARY; PRT; 455 AA.
AC Q9ZVY5;

```



```

RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,
RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA SCHUELLER C., CHALWATZIS N.,
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97339; CAB10327.1; -.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 479 AA; 53913 MW; 2ED16CCC CRC32;

Query Match      28.5%; Score 667; DB 10; Length 479;
Best Local Similarity 33.8%; Pred. No. 3.4e-47;
Matches 155; Conservative 88; Mismatches 176; Indels 40; Gaps
13;

QY 6 VLLATFPAQGHINPALQFAKRLKAGTDTFTFTSVYAW---RRMANTASAAAGNPPGLDF 62
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 VMLVSPGQGHVNPRLRLGLKLIASKGLLVTFVTEKPMGKKMRQANKIQDGVLPVGLGF 68
QY 63 VA---FSDGDDGLKPGDGKRYMSEMKARGSEALRNLL--LNDDDTFVYVSHLPAWAA 117
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 59 IRFEFFSDGFADDDKRFDPDAPRPLEAVGQEKALNPKRYNKEPVTCLINNAFVFWC 128
QY 118 EVARLSHVPATLLWVEPATVLCIYHFYNGYA-----DEIDAGSNEIQLPRLPSLEQRL 172
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 DVAELHPSAVLVQSCAGLTAYYYYHRLVKFTKTEPDT---SVEIPCLLKKHDEI 185
QY 173 PTFLLPATP-ERFLRMKKELETLDCEERAKVLVNTFDALPDALTAIDRYL---ELIGI 227
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 PSFLHSPSYTAFGDIIDLOLKFENHKSFYLFIDTFRELEKINDHMSQLCPQAIISPV 245
QY 228 GPLPSAFLEDGEDPSETSYGGDLFEKSENNVEMLNPKSSVYVVSFGSVLRFPPKAQM 287
DB 246 GPLPKMAQTLSDDVK-----GDISEPA--SDCMEWLDSEPSVYVVSFGSVLRFPPKAQM 298
QY 288 EEIGKGLACGRPLMIRQKNDGEEEBEELSCIGELKMKKIVSMCSQLEVLNHP 347
DB 299 EEIAGVLSGLSLVWVRPMEGTVEFVHLPR-----ELEEGKIVENCQPERVLNHP 353
QY 348 ALGCFVTHCGWNSAVESLSGIPVAVPWFQDTTNAKLIEDAGTGVYVRMNGGVDG 407
DB 354 AIACFLSHCGWNSMTMEALTAGVPVVCVFPQMGDQVTDVYLDVFKTVGLRGLR---GAAEE 410
QY 408 CEIERCV--EMVMDG--GDKTLVRENAIKKTKLARQAM 442
DB 411 MIVSREVVAEKLLEATVGERKAVELRENARRWKAEEAAV 449

RESULT 11
O22822 PRELIMINARY; PRT; 449 AA.
AC O22822;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE ISOLOG.
GN F18019.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euryhalophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA VENTER J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AC002333; AAB64024.1; -.
DR MENDEL; 26507; Arath;3145;26507.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 449 AA; 50771 MW; 8B0DCE0A CRC32;

Query Match      28.1%; Score 657.5; DB 10; Length 449;
Best Local Similarity 33.9%; Pred. No. 1.9e-46;
Matches 156; Conservative 92; Mismatches 157; Indels 55; Gaps
16;

QY 3 RRRVLLATFPAQGHINPALQFAKRLKAGTDTFTFTSVYAWRRMANTASAAAGNPPGLDF 62
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 RGHVLAVPYPTQGHITPFCRCRLHFKGLTKTLALTTFVF---NSINPDLSP--ISI 58
QY 63 VAFSDGDDGLKPGDGK-KRYMSEMKARGSEALRNLL-----LNDDDTFVYVSHLPAWAA 117
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 59 ATISDGYDHGGFETADSIDYLDKFTSGSKTIADIQKHQTSNPITCIYVDAFLPMAL 118
QY 118 EVARLSHVPATLLWVEPATVLCIYHFYNGYADEIDAGSNEIQLPRLPSLEQRLSPTF-- 175
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 DVAREFGIVATPPTPTCAVNVY-----YLSYINNGSLQLPIEELFLEQLDLPSPFS 172
QY 176 LLPATPFRFLMKKELETLDGEEKAK-VLVNTFDALPDALTAIDRYEL-----IGI 227
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 VSGSYPAIFEMVQQQFINF-----EKADFVLVNSFOELE-----LHENELSKACPVLT 222
QY 228 GPLPSAFLEDGEDPSETSYGGDLFEKSENNVEMLNPKSSVYVVSFGSVLRFPPKAQM 287
DB 223 GPTIPSYLDQRIKSDGTGYDLNLFESKDDSFCLNMLDTRPGSVYVVSFGSMAQLTNVQM 282
QY 288 EEIGKGLACGRPLMIRQKNDGEEEBEELSCIGELKMKK-----IVSMCSQLEV 343
DB 283 EELASAV--SNFSLVWVR-----SSEELKPS-GFLETVNKESLVKWSPLQV 330
QY 344 LAHPALGCFVTHCGWNSAVESLSGIPVAVPWFQDTTNAKLIEDAGTGVYVRMNGG 403
DB 331 LSNKAIGCFLTHCGWNSMTMEALTGFVPMVAMPQWTDQPNNAKYIQDVNKGAGVYVKTES 390
QY 404 GV-DCETERCVEMVMDGDKTLVRENAIKKTKLARQAM 442
DB 391 GIAKREEIEFSIKVME-GERSKMKKNVKKWRDLAVKSL 429

RESULT 12
O22820 PRELIMINARY; PRT; 449 AA.
AC O22820;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE ISOLOG.
GN F18019.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euryhalophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA VENTER J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002333; AAB64022.1; -.
DR MENDEL; 26506; Arath;3145;26506.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 449 AA; 50331 MW; FB3CC5C8 CRC32;

```



```
Query Match 27.6%; Score 644.5; DB 10; Length 449;
Best Local Similarity 34.8%; Pred. No. 2.2e-45;
Matches 159; Conservative 82; Mismatches 169; Indels 47; Gaps 17;

QY 2 VRRVLLATFPAGHINPALQAKRLKAGTDVTFVSVAVRRMANTASAAAGNPPGLD 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 MRGHVLAVPPSQGHITPQRCKRLHSGKFKTHLTITFP-----NTHLDPSSP--IS 57
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 FVAFSDGYD--GLKPGDGKRYMSEMKARGSEALRNLL-----LNNDVTFVYVSHLFA 116
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 IATISDGYDGGFSSAGSPEYQLQFKTGFSGKTADITIRKHQSTDNIPITCIVYDSPMPWA 117
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 117 AEVARLSHVTPTALLMAYEPATVLCIYHFYNGYADEIDAGSNEIQPLRPSLEQSRSLP 176
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 LDLANDFGLAAPFFTQSCAV-----NYIN-YLSYINNGSULTPIKDLPLELQDLPTFV 171
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 177 LPATPFRFLMKKEK-LETLDGEERAK-VLVNTP---DALEPDALTAIDRYELIGPLI 231
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 172 ---TPTGSHLAYFEMVLOQFTNFRADQFVLVNSFHDLDLHEEELLSKV---CPVLTIGPTV 226
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 232 PS AFLDGEOPSETSYGGDLFEKSEENNCVEMLSKPKSSVYVSGSVLRFPPKAQMEETG 291
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 227 PSMVLDQIKSDNDYDLNLFOLKEAALCTDMLDKRPGESVYVYAFGSMKLSSEQMEETA 286
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 292 KGLACGRFLMIR---BQKNDGEEEBEELSCIGELKMKKIVSQCQVLELAHPA 348
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 287 SAT--SNFSYLMVVRASEESKLPFGFLETVDKDKSL-----VLKWSPLQVLSNKA 335
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 349 LGCFTVTHCGWNSAVESLSGIGPVVAVPQFQDTTNAKLIEDAWGTGVVRNMGEGVDGC 408
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 336 IGCFTMTHCGWNSMTEGLSLGVPVAVPQWTDOPMNAKYIQDVWVGVVRVKAESGI--C 393
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 409 ---EIERCEVEMVMDGDKTKLVRENAIKWTKLARQAM 442
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 394 KREEIEFSIKEVNE-GEKSKEMKENAGKWRDLAVKSL 429
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
O48676 PRELIMINARY; PRT; 460 AA.
AC O48676;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DE F316.2 PROTEIN.
GN F316.2
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
RA AU M., ARAGJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,
RA QUI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VIJOTSKAYA V.S.,
RA YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002396; AAC00570.1;
DR MENDEL; 27887; Arath; 3145; 27887.
DR PROSITE; PS00375; UDRPGT; 1.
DR PFAM; PF00201; UDRPGT; 1.
SQ SEQUENCE 460 AA; 51002 MW; 02F08C19 CRC32;

Query Match 27.4%; Score 639.5; DB 10; Length 460;
Best Local Similarity 34.8%; Pred. No. 6e-45;
Matches 162; Conservative 66; Mismatches 181; Indels 57; Gaps 15;

QY 2 VRRVLLATFPAGHINPALQAKRLKAGTDVTFVSVAVRRMANTASAAAGNPPGLD 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Query Match 26.6%; Score 621.5; DB 10; Length 456;
Best Local Similarity 34.7%; Pred. No. 1.8e-43;
Matches 156; Conservative 86; Mismatches 171; Indels 37; Gaps 14;

QY 6 VLLATFPAGHINPALQAKRLKAGTDVTF-FTSVAVRRMANTASAAAGNPPGLDFA 64
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 11 VLMVTLTFFOCHINPMLKLAHLKLSLKNLHINLATIESARDLLSTVEKPR--YVDLVF 67
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 65 FSDGYDGLKPGDGKRYMSEMKARGSEALRNLL-----LNNDVTFVYVSHLFAAEVAVRL 122
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 68 FSDGL-----PKEDRAPETLLKSLNKGVMNLSKIIIEKRYSCIISPPFPWPAVAS 122
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 123 SHVPTALLWTEPATVLCIYHFYF---NGYADEIDAGSNEIQPLRPSLEQSRSLPFL 179
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
DB 123 HNSICALLWQACGAYSVYRYKYNISPPDLEDL-NQTVLPALPLLEVRDLSPMLPS 181
QY 180 TPERFLMMKEKLETLDEGEKAKVLYNTDPALEPDALTAI-DRYELIGIGLIPSAFLDG 238
DB 182 GGAHFYNLMAEFADCL--RVYKWLNVNSFVESEIIESMADLKVPIDIGPLV-SPFLIG 238
QY 239 EDPSETS YGGDLFEKSENNCVMENLNSKPKSSVYVYSGVSLRFPKAOHEEIGKGLLACG 298
DB 239 DGEETLDGKMLDCKSDCCNEWLKQARSSVYISFGSNLETLLENQVETIAKLNKRG 298
QY 299 RFLWMIREQKNDGEEEBEELSICIGELKKMGK--IVSWCSQLEVLALHPALGCFVTHC 356
DB 299 LPFLWVIR-----PKEKAQNAVQLQEMVKEGCGVLEWSPQEKILSHEAISCFVTHC 350
QY 357 GWSAVESLSGIPVAVPQWFQDTTNAKLIEDANCTGVYVRMNEGGVDG---CEIER 412
DB 351 GWNSTWETVAVGVPAVPSWTQPDARLLVDVFGVGRM---NDSVDELKVEEVER 407
QY 413 CVMVMGDDGDKTLVRENAIKWKTLARQAM 442
DB 408 CIEAVTEGPAAVD-IRRRRAELKRVARLAL 436

RESULT 15
O22183 PRELIMINARY; PRT; 438 AA.
AC O22183;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE PUTATIVE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
GN T20D15.12
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DDJ databases.
DR EMBL; AC002391; AAB87106.1; -
DR MENDEL; 26262; Arath; 3145; 26262.
DR PROSITE; PS00375; UDPGT; 1.
DR PAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 438 AA; 48717 MW; A8P76990 CRC32;
```

Query Match 26.3%; Score 613.5; DB 10; Length 438;
Best Local Similarity 34.8%; Pred. No. 7.9e-43;
Matches 154; Conservative 86; Mismatches 158; Indels 45; Gaps 18;

```
QY 14 QGHTNPALQPAKRLKAGTDVTFETSVYARRMANTASAAAGNPPGLDFVAFSDGY--DD 71
DB 7 QOHLNPMUKFAKHLARTNLHTLATTEQARDLLSSTADEP--HRP-VDLAFSGLPKDD 63
QY 72 GLKPGGCKRYMSEMKARGSEALRNLLNN--DDVTFVYVYSHLFAAAEVARLSHVPTAL 129
DB 64 PRDPTLAK-----SLKKDGAKNLSKIEEKRFDCIISVPFT---PWVPAVAHAHNIPCAI 116
QY 130 LWVEPATVLCIYHYFV---NGYADEIDAGSNEIQLPRLPSLEQRSLSPTFLPATPERFL 186
DB 117 LWIQACGAFSVYRYRYKYNISPPDLEDL-NQTVLPALPLLEVRDLSPMLPSQGANVNT 175
QY 187 MMKEKLETLDEGEKAKVLYNTDPALEPDALTAI-DRYELIGIGLIPSAFLDGDPSETS 245
DB 176 LMAEFADCL--KDVYKWLNVNSFVESEIIESMSDLKPIPIGPLV-SPFLIGNDEKTL 232
```

```
QY 246 YGGDLFEKSENNCVMENLNSKPKSSVYVYSGVSLRFPKAOHEEIGKGLLACGRPFLWMI 305
DB 233 ---DMWK--VDDYCMEWLDKQARSSVYISFGSILKSLENQVETIATALKNRGVPEFLWVI 287
QY 306 REQKNDGEEEBEELSICIGELKKMGK--IVSWCSQLEVLALHPALGCFVTHCGWNSAVE 363
DB 288 R-----PKEKGENVQLQEMVKEGCGVYVTEWGOQEKILSHMAISCFITHCGWNSTIE 339
QY 364 SLSGIPVAVPQWFQDTTNAKLIEDANCTGVYVRMNEGGVDG---CEIERCVMWMD 419
DB 340 TVVTGVPVAVPTPIDQPLDARLLVDVFGVGRM---NDAIDGELKVAEVEERCIEAVTE 396
QY 420 GGDKTKLVRENAIKWKTLARQAM 442
DB 397 GPAAADM-RRRATLKLKAARSAM 418

Search completed: August 1, 2000, 21:31:35
Job time: 6064 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 1, 2000, 19:51:30 ; Search time 59,57 Seconds
(without alignments)
113,997 Million cell updates/sec

Title: US-09-147-955-4

Perfect score: 2337
Sequence: 1 MYRRVLLATFPAQGHINP.....TKLVRENAIKWKLARQAMG 443

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/PCUS_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	707	30.3	471	2	US-08-466-583-2
2	707	30.3	471	4	PCT-US95-07820-2
3	383.5	16.4	131	2	US-08-466-583-5
4	383.5	16.4	131	4	PCT-US95-07820-5
5	375.5	16.1	471	3	US-09-106-464-2
6	368	15.7	488	2	US-08-797-228-2
7	244.5	10.5	131	2	US-08-466-583-6
8	244.5	10.5	131	4	PCT-US95-07820-6
9	219	9.4	63	2	US-08-466-583-8
10	219	9.4	63	4	PCT-US95-07820-8
11	160	6.8	529	4	PCT-US92-00282-7
12	155	6.6	531	4	PCT-US92-00282-6
13	146.5	6.3	534	4	PCT-US92-00282-4
14	145.5	6.2	533	4	PCT-US92-00282-5
15	143	6.1	533	4	PCT-US92-00282-3
16	121.5	5.2	506	5	5180581-2
17	119	5.1	56	2	US-08-466-583-4
18	119	5.1	56	4	PCT-US95-07820-4
19	107	4.6	58	2	US-08-466-583-9
20	107	4.6	58	4	PCT-US95-07820-9
21	104	4.5	74	4	PCT-US92-00282-24
22	103.5	4.4	399	1	US-08-096-623A-20
23	101	4.3	52	2	US-08-466-583-7
24	101	4.3	52	4	PCT-US95-07820-7
25	99	4.2	414	2	US-08-750-524-1
26	90	3.9	1271	1	US-08-095-734-2
27	90	3.9	1271	2	US-08-444-623-2
28	90	3.9	1271	3	US-08-471-869-2

29	90	3.9	1271	4	PCT-US94-08267-2	Sequence 2, Appli
30	88.5	3.8	421	2	US-08-576-626A-53	Sequence 53, Appl
31	88.5	3.8	834	1	US-07-977-434-10	Sequence 10, Appl
32	88.5	3.8	834	1	US-08-073-384C-6	Sequence 6, Appl
33	88.5	3.8	834	1	US-08-254-359A-6	Sequence 6, Appl
34	88.5	3.8	834	1	US-08-384-490-31	Sequence 31, Appl
35	88.5	3.8	834	1	US-08-483-043-6	Sequence 6, Appl
36	88.5	3.8	834	1	US-08-459-383-31	Sequence 31, Appl
37	88.5	3.8	834	1	US-08-458-819-10	Sequence 10, Appl
38	88.5	3.8	834	1	US-08-481-238-6	Sequence 6, Appl
39	88.5	3.8	834	2	US-08-484-956-6	Sequence 6, Appl
40	88.5	3.8	834	2	US-08-757-653-6	Sequence 6, Appl
41	88.5	3.8	834	2	US-08-599-491-6	Sequence 6, Appl
42	88.5	3.8	834	2	US-08-756-386-6	Sequence 6, Appl
43	88.5	3.8	834	2	US-08-823-516-6	Sequence 6, Appl
44	88.5	3.8	834	2	US-08-682-853A-6	Sequence 6, Appl
45	88.5	3.8	834	3		

ALIGNMENTS

RESULT 1
US-08-466-583-2
; Sequence 2, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szczyglowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-583-2

Query Match 30.3%; Score 707; DB 2; Length 471;
Best Local Similarity 37.4%; Pred. No. 5.6e-66;
Matches 174; Conservative 70; Mismatches 171; Indels 50; Gaps 16;

QY 6 VLLATFPAQHINPALQFAKRLKAGTDVTFSTVYARRMANTASAAAGNPPGLDFVAF 65
Db 5 VLVVPPFGQGHMPVQFAKRLKAGTATLVTT-----RFIQTADVDA-HPAMVE--AI 57
QY 66 SDGYDDGLKPGDG-KRYMSEMKARGSEALRNLL-----LNDDVTFVYVYSHLFAWAEVA 120
Db 58 SDGHDEGGFASAGVAYEYLEKQAAASASLASLVEARASSADAFTCVVYDSYEDWVLPVA 117
QY 121 RLSHVPTALLWPEPATVLCIYHFYNG-----YADEIDAGSNEIQLPR----LPSLE 168
Db 118 RMGLPAVPFSTQSCAVSYVYHFSQGLAVPPGAAADGSGGAGAAALSEAFGLGPEME 177
QY 169 QRSUPTFLLPATPERFLMKKELETLDGEERAKVLVNTFDALPDAITADRY-ELIGI 227
Db 178 RSELPFVFDHGPVPTTAMQAIQFAHAGKDD-WVLFNSFEETEVLGTLKYLKARAI 236
QY 228 GPLP-----SAPLGDGEPSETS YGGDLFEKSEENNCVENLNKPKSSVYVYSGSVLR 281
Db 237 GPCVPLPTAGTAGANG----RITYGANLVK--PEDACTKWLDTKPDRAVYVSGSLAS 290
QY 282 FPKAQMBEIGKGLLACGRPFLLMIREQKNDGCEEEBEEELSCIGELKMG--KIVSWCS 339
Db 291 LGNAQKEELARGLLAAGKPFLLWVR-----ASDEHQVPRYLLAEATATGAAMVVPWCP 343
QY 340 QLEVLHPALGCFVTHCGWNSAVESLSCGIPVAVPQNFDTNNAKLIEDANGTGVVRM 399
Db 344 QLDVLAHPAVGCFVTHCGWNSLEALSFGVPMVAMALWTDQPTNARNVELANGAGVRARR 403
QY 400 NEGCGV--DGCEIERCVEMVMDGDKTKLVRENAIKWKTARQAM 442
Db 404 DAGAGVFLRG-EVERCVRAVMDGGEAASAARKAAGEWRDRARA 447

RESULT 2

PCT-US95-07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07820-2

Query Match 30.3%; Score 707; DB 4; Length 471;
Best Local Similarity 37.4%; Pred No. 5, 6e-66;
Matches 174; Conservative 70; Mismatches 171; Indels 50; Gaps 16;

QY 6 VLLATFPAQHINPALQFAKRLKAGTDVTFSTVYARRMANTASAAAGNPPGLDFVAF 65
Db 5 VLVVPPFGQGHMPVQFAKRLKAGTATLVTT-----RFIQTADVDA-HPAMVE--AI 57
QY 66 SDGYDDGLKPGDG-KRYMSEMKARGSEALRNLL-----LNDDVTFVYVYSHLFAWAEVA 120
Db 58 SDGHDEGGFASAGVAYEYLEKQAAASASLASLVEARASSADAFTCVVYDSYEDWVLPVA 117
QY 121 RLSHVPTALLWPEPATVLCIYHFYNG-----YADEIDAGSNEIQLPR----LPSLE 168
Db 118 RMGLPAVPFSTQSCAVSYVYHFSQGLAVPPGAAADGSGGAGAAALSEAFGLGPEME 177
QY 169 QRSUPTFLLPATPERFLMKKELETLDGEERAKVLVNTFDALPDAITADRY-ELIGI 227
Db 178 RSELPFVFDHGPVPTTAMQAIQFAHAGKDD-WVLFNSFEETEVLGTLKYLKARAI 236
QY 228 GPLP-----SAPLGDGEPSETS YGGDLFEKSEENNCVENLNKPKSSVYVYSGSVLR 281
Db 237 GPCVPLPTAGTAGANG----RITYGANLVK--PEDACTKWLDTKPDRAVYVSGSLAS 290
QY 282 FPKAQMBEIGKGLLACGRPFLLMIREQKNDGCEEEBEEELSCIGELKMG--KIVSWCS 339
Db 291 LGNAQKEELARGLLAAGKPFLLWVR-----ASDEHQVPRYLLAEATATGAAMVVPWCP 343
QY 340 QLEVLHPALGCFVTHCGWNSAVESLSCGIPVAVPQNFDTNNAKLIEDANGTGVVRM 399
Db 344 QLDVLAHPAVGCFVTHCGWNSLEALSFGVPMVAMALWTDQPTNARNVELANGAGVRARR 403
QY 400 NEGCGV--DGCEIERCVEMVMDGDKTKLVRENAIKWKTARQAM 442
Db 404 DAGAGVFLRG-EVERCVRAVMDGGEAASAARKAAGEWRDRARA 447

RESULT 3
US-08-466-583-5
; Sequence 5, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; PCT-US95-07820-5
;
Query Match 16.4%; Score 383.5; DB 2; Length 131;
Best Local Similarity 50.7%; Pred. No. 8.6e-33;
Matches 70; Conservative 22; Mismatches 37; Indels 9; Gaps 2;
US-08-466-583-5
QY 259 CVELNSKPKSSVVVSGSVLRFPPKAQMEIEGKGILLACGRPFLMWIREQNNDGEREE 318
Db 1 CTWKLDTPDRSVAYVSFGSLASGNQAQEELARGLLAAAGKPFLLWVR-----ASDEH 53
QY 319 EEELSICIGELKKMG--KIVSWCSOLEVLHAHPALGCFTVTHCGWNLSAVESLSGPIPVAVPQ 376
Db 54 QVPRYLLAEATATGAAMVVPWCPLDVLHAHPAVGCFVTHCGWNSTLEALSFGVPVMVAL 113
QY 377 WFDQTTNAKLIEDAWGTG 394
Db 114 WTDQPTNARNVELANGAG 131
;
RESULT 4
PCT-US95-07820-5
; Sequence 5, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jedrzej B.
; APPLICANT: Szczylowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
;
Query Match 16.4%; Score 383.5; DB 2; Length 131;
Best Local Similarity 50.7%; Pred. No. 8.6e-33;
Matches 70; Conservative 22; Mismatches 37; Indels 9; Gaps 2;
US-08-466-583-5
QY 259 CVELNSKPKSSVVVSGSVLRFPPKAQMEIEGKGILLACGRPFLMWIREQNNDGEREE 318
Db 1 CTWKLDTPDRSVAYVSFGSLASGNQAQEELARGLLAAAGKPFLLWVR-----ASDEH 53
QY 319 EEELSICIGELKKMG--KIVSWCSOLEVLHAHPALGCFTVTHCGWNLSAVESLSGPIPVAVPQ 376
Db 54 QVPRYLLAEATATGAAMVVPWCPLDVLHAHPAVGCFVTHCGWNSTLEALSFGVPVMVAL 113
QY 377 WFDQTTNAKLIEDAWGTG 394
Db 114 WTDQPTNARNVELANGAG 131
;
RESULT 5
US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 601145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Changas, Gurdev S.
; APPLICANT: Kual, Jian-Ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc:Fatty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,464
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,554
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Spector, Eric S.
; REGISTRATION NUMBER: 22495
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-415-1500
; TELEFAX: 703-415-1508
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-09-106-464-2

Query Match 16.1%; Score 375.5; DB 3; Length 471;
Best Local Similarity 27.5%; Pred. No. 4.7e-31;
Matches 137; Conservative 65; Mismatches 173; Indels 123; Gaps 22;

Qy 11 FP--AQGHINPALQFAKRLKAGTDTFTT-----SVYAWRMAN-----TAS 51
Db 10 FPMQAQGHMPTLDMAKLVACRGVKATITTPLNESVFSKAERKKGIGIDIRLKKFP 69
Qy 52 AAAGNPP---GLDFVAFSDGDKLPGGDKRYMSEKARGSEALNLLNDDVTVF 107
Db 70 AKENDLPEDCERLDLPVSDDKLPFLKAAAMKDEFEE-----LIGECRDPCL 117
Qy 108 VYSHLFAWAEVAVRLSHVPTALLWVEPATVLCIYHFYFNGVADIDAGSNEIQLPRLPSL 167
Db 118 VSDMFLPWTDSAAKSPKPRVPHGYSYFALCV-----GDSI----- 154
Qy 168 BQSLP-----TFLPATPERFLMKKE--KLETLD-----GEEKAK-- 202
Db 155 -RRNPKPKNVSSDTFTVVPDFPHEIRLRTQLSPFEQSDDETGMPIKAVRESDAKY 213
Qy 203 -VLNTDALEPALTAIDRY-ELIG-----IGPLPSAFLDGDPSETS YGGDLFEKS- 254
Db 214 GVINSFYELED---YVEHYKVGKKNWALGPL-----SLCNRD---IEYKAERGRKSS 263
Qy 255 -EENNVEWLNKPKSVVYVSGSVLRFPKQMBEEIGKLLACGRPFELMIRQKND-- 311
Db 264 IDEHACLKWLDSKSSIVYVFCGSTADPTTAQMOELAMGLEASQODFIWIRTGNDWL 323
Qy 312 -DGEEREEELSCIGELKMKGRIV-SWCSQLEVLHAPALGCFVTHCGWNSAVESLSCGI 369
Db 324 PEGFEERTKEK-----GLIIGWAPQVLIIDHEAIGAFVTHCGWNSTLEGISAGV 373
Qy 370 PVAVPQWFODTNNAKIEDWGTGVYV-----RNNEGVDGGEIERCEVEMVMDGDKT 424
Db 374 PMLTWPFVFAEQFTNKLVTEVMSRSGAGVGSKQMKRTASEGVKREAIKAKIRVN-ASEET 432
Qy 425 KLVRENAIKWKTARQAM 442
Db 433 EGFRSRAKEYKEMAREAI 450

RESULT 6
US-08-797-226-2
; Sequence 2, Application US/08797226
; Patent No. 5959180
; GENERAL INFORMATION:
; APPLICANT: MOEHS, CHARLES P
; APPLICANT: ALLEN, PAUL V
; APPLICANT: ROCKHOLD, DAVID R
; APPLICANT: STAPLETON, ANDREW
; APPLICANT: GARBARINO, JOAN E
; APPLICANT: FRIEDMAN, MENDEL
; APPLICANT: BELKNAP, WILLIAM R
; TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
; TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
; TITLE OF INVENTION: GLYCOLALDEHYDES IN SOLANACEOUS PLANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NANCY J. PARSONS
; STREET: 800 BUCHANAN ST.
; CITY: ALBANY
; STATE: CA
; COUNTRY: USA
; ZIP: 94710
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/797,226
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PARSONS, NANCY J
; REGISTRATION NUMBER: 40,364
; REFERENCE/DOCKET NUMBER: 0011.97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 559-5731
; TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 488 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-797-226-2

Query Match 15.7%; Score 368; DB 2; Length 488;
Best Local Similarity 23.1%; Pred. No. 3.1e-30;
Matches 117; Conservative 93; Mismatches 172; Indels 124; Gaps 20;

Qy 6 VLLATFPAQGHINPALQFAKRLKAGTDTFTT-----SVYAWRMAN----- 49
Db 13 VLFPLFSLAGHFLPVNAARLFASRGVKATITTPHALLFRSTIDDDVRISGPIAIVT 72
Qy 50 ---ASAAAGNPPGLDFVAFSDGDKLPGGDKRY--MSEMKARGSEALNLLNDDV 104
Db 73 IKFPSAEVGLPEGI-----ESFNSATSPMPHKIFYALSLLQKPMEDKIRELRPD--- 122
Qy 105 TFVYVSHL-FAWAEVAVRLSHVPTALLWVEPATVLCIYHFYFNGVADIDAGSNEIOLP- 162
Db 123 --CFPSDMYFPTVTDIADHLHPRILYNLSAYWCYSIMH-----NLYVYRPH 167
Qy 163 RLPSEORSPLTFLPATPERFLMKKEKLETL-----DGEKAKVL 204
Db 168 KQNLDESQ--SFVYVGLPDEIKFKLSQTLDDLKRSDDQKTVFDELLEQVEDSEERSYGI 225
Qy 205 V-NTFDALPDALTADRYELI-----GIGPLPSAFLDGDPSETS YGGDLFEK--S 254
Db 226 VHDTFYELEP--AYVDYQYKAKPKCHWFGPL-----SHFASKIRSKELIS 269
Qy 255 EENN---CWEWLNKPKSVVYVSGSVLRFPKQMBEEIGKLLACGRPFELMIRQKND 311
Db 270 EHNNEIVIDLWLNQAKPKSVLYVSGSMARFPESQLNETAQALDASNVYFFIEVLR----- 324
Qy 312 DGEEREEELSCIGELKMKGK---IVSWCSQLEVLHAPALGCFVTHCGWNSAVESLSC 367
Db 325 ---PNEETASWLPVGNLEDKTKGLYIKGVFPQLIMEHSATGGTGCCTNSVLEAITP 381
Qy 368 GIPVAVPQWFODTNNAKLE-----DAGTGVYVRNMEGGVGDGEIERCEV-- 415
Db 382 GVPMTWPLVADQFYNEKVVVEVGRGLGKIGIDVWNEGIBI---TGPVIESAKIREAIERL 438
Qy 416 VMVMDGGDKTKLVRENAIKWKTARQ 441
Db 439 MINSGBEINIRDRVWMSKMAQNA 464

RESULT 7
US-08-466-583-6
; Sequence 6, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szareszen, Jędrzej B.
; APPLICANT: Szczeglowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.


```

RESULT 10
PCT-US95-07820-8
: Sequence 8, Application PC/TUS9507820
: GENERAL INFORMATION:
: APPLICANT: Bandurski, Robert S.
: APPLICANT: Szerszen, Jedrzej B.
: APPLICANT: Szczeglowski, Krzysztof
: TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
: TITLE OF INVENTION: and Plant Growth.
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee & Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/07820
: FILING DATE: 19-JUN-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/265,427
: FILING DATE: 24-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ferber, Donna M.
: REGISTRATION NUMBER: 33,878
: REFERENCE/DOCKET NUMBER: 11-94B PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 303/499-8080
: TELEFAX: 303/499-8089
: TELEX: 49617824
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 63 amino acids

```

122	QY	LSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNETQLP-RUPSLUEQRSLSFTFLLPAT	180
177	DB	LEHTGQS-----PSPVSYVPRFY-----TKFSDHMTFFORLANTIANILENYLYHCL	223
181	QY	PERFLMMKESKLETLDGEERAKVLVNTFOALEPFDALTAIDRYELI--GIGPLIPS-AFLD	237
224	DB	YSKVEIILASDLL-----KRDY---SLPALHONSLWL--RYDFVFGVPRVPMNMFIF	272

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:19 ; Search time 47.63 Seconds
(without alignments)
220.301 Million cell updates/sec

Title: US-09-147-955-4

Perfect score: 2337

Sequence: 1 MVRRLVLLATFPAQGHINPA.....TKLVRENAIKWKTARQAMG.443

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2337	100.0	443	1 W92948	WO9905287 Seq ID 8
2	2223	95.1	460	1 W92947	WO9905287 Seq ID 7
3	1583.5	67.8	461	1 W92949	WO9905287 Seq ID 9
4	1412.5	60.4	478	1 W92950	WO9905287 Seq ID 1
5	1307.5	55.9	468	1 W92952	WO9905287 Seq ID 1
6	707	30.3	471	1 W03756	Maize UDP-glucose
7	659	28.2	466	1 W09825	UDP-glucose:thiohy
8	384	16.4	433	1 W13406	Solanum melongena
9	378.5	16.2	471	1 W98009	Medium-chain UDP-g
10	374.5	16.0	470	1 W47372	Glucosyl transfera
11	368	15.7	488	1 W64369	Potato solanidine
12	328.5	14.1	471	1 R49245	Anthocyanidin-3-gl
13	324	13.9	473	1 W56451	UDP-glucose:flavon
14	135.5	5.8	534	1 R26154	HUG-Br2. Isolated
15	134	5.7	533	1 R26153	HUG-Br1. Isolated
16	125	5.3	530	1 W47126	Uridine diphospho-
17	124	5.3	515	1 W67250	Ecdysteroid UDP-gl
18	121.5	5.2	506	1 R10429	Ecdysteroid UDP-gl
19	110.5	4.7	399	1 R13989	zeaxanthin glycosy
20	104	4.5	74	1 R30165	UGT1 exon 4 produc
21	103	4.4	1198	1 W64384	S. cerevisiae L947
22	101	4.3	473	1 W64392	S. cerevisiae L947
23	99	4.2	414	1 R93982	Saccharothrix aro
24	96.5	4.1	421	1 W99388	S. erythraea desosa
25	96	4.1	431	1 Y00181	Enterococcus faeca
26	96	4.1	474	1 Y00180	Enterococcus faeca
27	94	4.0	508	1 W35712	Torenia flavonoid
28	91.5	3.9	594	1 R34936	CENP-B. Human cent
29	91.5	3.9	1012	1 W63671	Polypeptide having
30	90	3.9	931	1 Y00872	Original S. tubero
31	89.5	3.8	725	1 W98106	Human elongation f
32	88.5	3.8	421	1 W19735	Sugar biosynthesis
33	88.5	3.8	758	1 R23165	Mutant thermostabl

ALIGNMENTS

RESULT 1

W92948

ID W92948 standard; Protein; 443 AA.

AC W92948;

DT 14-MAY-1999 (first entry)

DE WO9905287 Seq ID 8.

KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety: colour.

OS Perilla frutescens.

PN WO9905287-Al.

PD 04-FEB-1999.

PF 16-JUL-1998; J03199.

PR 25-JUL-1997; JP-200571.

PA (SUNR) SUNTORY LTD.

PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,

PI Yamazaki M.

DR N-PSDB; X02827.

DR N-PSDB; X02827.

PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be

PT used to transform plants for improvement of plant coloration

PS Claim 2; Page 74-76; 89pp; Japanese.

CC This invention describes the isolation of plant proteins which have

CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used

CC in the production of plant varieties with improved colour, for example in

CC the production of flower varieties with desired coloration and cut

CC flowers harvested from them.

CC Sequence 443 AA;

Query Match

Best Local Similarity 100.0%; Score 2337; DB 1; Length 443;

Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRRLVLLATFPAQGHINPALQFAKRLKACTDVTFTTSYVYANRRMANTASAAAGNPPGL 60

Db 1 MVRRLVLLATFPAQGHINPALQFAKRLKACTDVTFTTSYVYANRRMANTASAAAGNPPGL 60

QY 61 DFVAFSDGVDGDLKPGDGKRYMSEKARGSEALRNLLNDDVTFVYVSHLFAWAAEVA 120

Db 61 DFVAFSDGVDGDLKPGDGKRYMSEKARGSEALRNLLNDDVTFVYVSHLFAWAAEVA 120

QY 121 RLSHVPALLWEPATVLCIHYFNGYADEIDAGSNEIQPLRPLSEQRSLFTLLPAT 180

Db 121 RLSHVPALLWEPATVLCIHYFNGYADEIDAGSNEIQPLRPLSEQRSLFTLLPAT 180

QY 181 PERFLMMEKLETLDGEEKAVLVNTFDALPDALTAIDRYELIGIGLIPSAFLDGED 240

Db 181 PERFLMMEKLETLDGEEKAVLVNTFDALPDALTAIDRYELIGIGLIPSAFLDGED 240

QY 241 PSTSTGGDLFEKSENNCVMNLNPKKSSVYVYVSGSVLRFPKQAMEETGKLLACGRP 300

Db 241 PSTSTGGDLFEKSENNCVMNLNPKKSSVYVYVSGSVLRFPKQAMEETGKLLACGRP 300

QY 301 FLWMIREQNDGEEEEEPEELSCIGELKMKGIKVCWCSOLEVLAHPALGCFVTHCGWNS 360

Db 301 FLWMIREQNDGEEEEEPEELSCIGELKMKGIKVCWCSOLEVLAHPALGCFVTHCGWNS 360

QY 361 AVESLSCGIPVAVPQWFQDTTNAKLIEDAGTGYVRVNMNEGGVGDGCEIERCVMYMDG 420
 Db 361 AVESLSCGIPVAVPQWFQDTTNAKLIEDAGTGYVRVNMNEGGVGDGCEIERCVMYMDG 420
 QY 421 GDKTKLVRENAIKWKTILARQAMG 443
 Db 421 GDKTKLVRENAIKWKTILARQAMG 443

RESULT 2
 ID W92947 standard; Protein: 460 AA.
 AC W92947;
 DT 14-MAY-1999 (first entry)
 DE W09905287 Seq ID 7
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
 OS *Perilla frutescens*.
 PN W09905287-A1.
 PD 04-FEB-1999.
 PF 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNR) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 PI Yamazaki M;
 DR WPI: 99-142940/12.
 DR N-PSDB: X02826.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 used to transform plants for improvement of plant coloration
 PS Claim 2; Page 71-73; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have
 flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
 in the production of plant varieties with improved colour, for example in
 the production of flower varieties with desired coloration and cut
 flowers harvested from them.
 CC flowers harvested from them.
 SQ Sequence 460 AA;

Query Match 95.1%; Score 2223; DB 1; Length 460;
 Best Local Similarity 95.7%; Pred. No. 3.7e-221;
 Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;
 QY 1 MVRRLVLLATFPAGQHINPALQFAKRLKAGTDVTFVSVYAWRRMANTASAAAGNPPGL 60
 Db 1 MVRRLVLLATFPAGQHINPALQFAKRLKAGTDVTFVSVYAWRRMANTASAAAGNPPGL 60
 QY 61 DFVAFSDGYDDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYVSHLFAWAEVA 120
 Db 61 DFVAFSDGYDDGLKPGDGKRYMSEMKARGSEALRNLLNNDVTFVYVSHLFAWAEVA 120
 QY 121 RLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNETQLPLRPSLEORSPLTFLLPAT 180
 Db 121 RESQVPSALLWVEPATVLCIYHFYFNGYADEIDAGSNETQLPLRPSLEORSPLTFLLPAT 180
 QY 181 PERFLRMKKEKLETLGDEEKAKVLVNTFDALPDALTADRYELIGIGLIPSLFDGED 240
 Db 181 PERFLRMKKEKLETLGDEEKAKVLVNTFDALPDALTADRYELIGIGLIPSLFDGED 240
 QY 241 PSETSYGGDLFEKSEENNCVWLNKPKSSVYVSGSVLPPKQAQMEIEGKGLLAGCRP 300
 Db 241 PSETSYGGDLFEKSEENNCVWLNKPKSSVYVSGSVLPPKQAQMEIEGKGLLAGCRP 300
 QY 301 FLWMIREQNDGDEEEELSCIGELKKMGKIYSWCSQLEVLAHPALGCFVTHCGWNS 360
 Db 301 FLWMIREQNDG--EEEEELSCIGELKKMGKIYSWCSQLEVLAHPALGCFVTHCGWNS 358
 QY 361 AVESLSCGIPVAVPQWFQDTTNAKLIEDAGTGYVRVNMNEGGVGDGCEIERCVMYMDG 420
 Db 359 AVESLSCGIPVAVPQWFQDTTNAKLIEDAGTGYVRVNMNEGGVGDGCEIERCVMYMDG 418
 QY 421 GDKTKLVRENAIKWKTILARQAMG 443
 Db 419 GEKSKLVRENAIKWKTILARQAMG 441

RESULT 3
 ID W92949 standard; Protein: 461 AA.
 AC W92949;
 DT 14-MAY-1999 (first entry)
 DE W09905287 Seq ID 9
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
 OS *Verbena hybrida*.
 PN W09905287-A1.
 PD 04-FEB-1999.
 PF 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNR) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 PI Yamazaki M;
 DR WPI: 99-142940/12.
 DR N-PSDB: X02828.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 used to transform plants for improvement of plant coloration
 PS Claim 2; Page 76-78; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have
 flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
 in the production of plant varieties with improved colour, for example in
 the production of flower varieties with desired coloration and cut
 flowers harvested from them.
 CC flowers harvested from them.
 SQ Sequence 461 AA;

Query Match 67.8%; Score 1583.5; DB 1; Length 461;
 Best Local Similarity 68.7%; Pred. No. 4.4e-155;
 Matches 310; Conservative 57; Mismatches 65; Indels 19; Gaps 7;
 QY 1 MVRRLVLLATFPAGQHINPALQFAKRLKAGTDVTFVSVYAWRRMANTASAAAGNPPGL 60
 Db 1 MSRAVLLATFPAGQHINPALQFAKRLANADIQVTFVSVYAWRRMNT---AAGSGLI 57
 QY 61 DFVAFSDGYDDGLKPGDGKRYMSEMKARGSEALRN--LLLNND---VTFVYVSHLFA 114
 Db 58 NFVSFSDGYDDGLQPGDGKRYMSEMKARGKALSDDTLAANNVDDQKSKITFVYVSHLFA 117
 QY 115 WAAEVARLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNETQLPRLPSLEORSPL 173
 Db 118 WAAKVARREFHLRSALLIEPATVLDIFIFNGYDEIDAGSDAHLPGGLFVLAQORDLP 177
 QY 174 TFLPATPEFRFLRMKKEKLETLGDEEKAKVLVNTFDALPDALTADRYELIGIGLIPPS 233
 Db 178 SFLLPSTHERFRSLMKKEKLETLGSEKPKVLYNSFDALPDALKALDKYEMTAIGLIPPS 237
 QY 234 AFLDGEDPSETSYGGDLFEK--SEENNCVWLNKPKSSVYVSGSVLPPKQAQMEIEGK 292
 Db 238 AFLDGPSPDSRSGDGLFEKGSNDQDCLWLTNPSSVYVSGSVVNTTKSQMEIEAR 297
 QY 293 GILLACGPRPFLWMIREQNDGDEEEELSCIGELKKMGKIYSWCSQLEVLAHPALGCF 352
 Db 298 GLLDGGPRPFLWVVR-----VNEGSEVLISCMEELKRVGKIYSWCSQLEVLTHPSLGP 350
 QY 353 VTHCGWNSAVESLSCGIPVAVPQWFQDTTNAKLIEDAGTGYVRVNMNEGGV--VDGGEIE 411
 Db 351 VTHCGWNTLESISFGVPMVAFQWFDQGTNAKLIEDVWRTGVRVRANEEGSVYVDGEIR 410
 QY 412 RCVEWYMDGGDKTKLVRENAIKWKTILARQAM 442
 Db 411 RCIEVMDGGEKSKRLRESAGKWKDLARKAM 441

RESULT 4
 ID W92950 standard; Protein: 478 AA.
 AC W92950;
 DT 14-MAY-1999 (first entry)
 DE W09905287 Seq ID 10.
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.

CC The present sequence is a UDP-glucose:fatty acid glucosyltransferase
 CC that transfers glucose to medium-chain length fatty acids to form
 CC 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses act as acyl donors
 CC in the esterification of glucose and in the further esterification
 CC of partially acylated glucose in reactions catalysed by glucose
 CC acyltransferases. They also act as acyl donors in the
 CC esterification of sucrose and in the further esterification of
 CC partially acylated sucrose in reactions catalysed by sucrose
 CC acyltransferases. A cDNA clone (see X24873) encoding the enzyme
 CC was isolated by immunoscreening a leaf trichome cDNA library
 CC prepared from an F1 population of a cross between wild tomato
 CC Lycopersicon pennellii (LA 716) and cultivated tomato
 CC Lycopersicon esculentum cv. New Yorker. The invention provides
 CC chain length-specific UDP-glucose:fatty acid glucosyltransferases
 CC and a method for their purification. The enzymes have specificity
 CC for transferring glucose to short, medium or long chain length
 CC fatty acids. Methods for preparing 1-O-acyl-a-D-glucoses, where
 CC the acyl group comprises 3-5, 6-13 or 14-22 C atoms, are claimed.
 SQ Sequence 471 AA;

Query Match 16.2%; Score 378.5; DB 1; Length 471;
 Best Local Similarity 28.1%; Pred. No. 1.5e-30;
 Matches 136; Conservative 69; Mismatches 184; Indels 95; Gaps 21;

QY 11 PP--AQGHINPALQFAKRLKAGTDVTF--SYAVRRMAN-----TAS 51
 DB 10 FPMQAQGHIMPTDMAKLVACRGVKATITPLNESVFSKATERNKHGIEIDIRLLKFP 69
 QY 52 AAAGNPP----GLDFVAFSDGDDGLKPGDGKRYMSEMKARGSEALRNLLNDDVTFV 107
 DB 70 AKENDLPDCERLDVPSDDKLPNFKAAAMMKDEFE-----LIGECRPDCL 117
 QY 108 VYSHLFAWAARVRLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNETQLPRLPSL 157
 DB 118 VSDMPLPMTTDSAAKFSIPRIVFHGTSYFALCVGHSIRNKPKNVSDTETVVPDPFH 177
 QY 158 EIQLPR--LPSLEORSLEPTLLPATPERFLAMKEKLETLDEGEKAKVLVNTFOALEPDA 215
 DB 178 EIRLTQTLSPEQSDTECMAP-----MIKAVRES--DAKSYGVIFNSFYELES- 226
 QY 216 LTAIDRY-ELIG-----IGLIPSAFLDGEDPSETSYGDLPEKS--BENNCVWLNKSP 267
 DB 227 --YVEHYTKVVGKKNWAIGPL-----SLCNRD---IEYKAERKRSIDHEACLKWLDSKK 277
 QY 268 KSSVYVSGSVLRFPKQAEMEEIGKGLLACGRPFLMIREKND---DGESEEEELSC 324
 DB 278 SSSIVYVCGSTADFTTAQMOELAMGLEASGQDFIWIIRTGNEWLPEGEERTKEK--- 334
 QY 325 IGEKKMGKIV-SNCSOLEVLAHPALGCFVTHCGWNSAVESLSCGIPVAVPQWFDQTN 383
 DB 335 -----GLIIRGWAPQVILDHEAIGAFVTHCGWNSLLEGISAGVPMLTWPFVFAOEFN 387
 QY 384 AKLIEDAMGTGVRY-----RMNEGGVYDGCIEIRCVEMVMDGGDKTLVRENAIKWKTLA 438
 DB 388 ERLVTEVMRSAGVSGKQWKRTASEGVKREAIKAKRYM-ASEETEGFRSRAKEYKEMA 446
 QY 439 ROAM 442
 DB 447 REAI 450

RESULT 10

W47172
 ID W47172 standard; Protein; 470 AA.
 AC W47172;
 DT 08-JUN-1998 (first entry)

DE Glucosyl transferase (Grase) protein encoded by TW11 gene.
 KW Glucosyl transferase; Grase; TW11; tomato; signalling pathway;
 KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;
 OS plant defence protein; plant response; tobacco; rice.
 OS Lycopersicon sp.
 PN W09745546-A1.

PD 04-DEC-1997.
 PF 30-MAY-1997; G01473.
 PR 31-MAY-1996; GB-011420.
 PA (UYVO-) UNIV YORK.
 PI Bowles DJ, Calvert CM, O'Donnell PJ, Roberts MR;
 DR WPI: 98-032653/03.
 DR N-PSDB; V17054.
 PT Tomato wound inducible (TW11) gene encoding glucosyl transferase -
 PT useful to develop products that alter signalling pathways in plants
 PT by altering of salicylic acid, jasmonic acid or ethylene
 PS Claim 2; Fig 3; 52pp; English.
 CC This is a glucosyl transferase (Grase) protein encoded by a wound
 CC inducible gene (TW11) isolated from wounded tomatoes. The TW11 gene
 CC encodes this Grase from amino acid position 5. The TW11 gene can be
 CC used to identify homologue Grase encoding genes isolated from tobacco
 CC and rice. A microbial host can be transfected or transformed with a
 CC vector containing the Grase encoding nucleic acids. The products can be
 CC used to interfere with Grase and therefore alter signalling pathways in
 CC plants, specifically tobacco, rice or tomato plants by altering levels
 CC of salicylic acid, jasmonic acid or ethylene. This can induce the
 CC production of plant defence proteins such as pathogenesis-related (PR)
 CC and proteinase inhibitor (PIN) proteins which regulate plant development
 CC (plant growth, reproduction and senescence) and improve plant response to
 CC pathogens. 470 AA;
 SQ Sequence 470 AA;

Query Match 16.0%; Score 374.5; DB 1; Length 470;

Best Local Similarity 27.4%; Pred. No. 3.8e-30;
 Matches 136; Conservative 64; Mismatches 177; Indels 119; Gaps 21;

QY 11 PP--AQGHINPALQFAKRLKAGTDVTF--SYAVRRMAN-----TAS 51
 DB 9 FPDQAQGHIMPTDMAANNVACRGVKATITPLNESVFSKATERNKHGIEIDIRLLKFP 68
 QY 52 AAAGNPP----GLDFVAFSDGDDGLKPGDGKRYMSEMKARGSEALRNLLNDDVTFV 107
 DB 69 AKENDLPDCERLDVPSDDKLPNFKAAAMMKDEFE-----LIGECRPDCL 116
 QY 108 VYSHLFAWAARVRLSHVPTALLWVEPATVLCIYHFYFNGYADEIDAGSNETQLPRLPSL 167
 DB 117 VSDMPLPMTTDSAAKFSIPRIVFHGTSYFALCV-----GDTI----- 153
 QY 168 EORSUP-----TELLPATPERFLMKME--KLETL-----GEEKAK-- 202
 DB 154 -RRNKPKNVNSDTEFVVPDLPHEIRLTQTLSPEQSDTEETGMPIKAVRESDAKSY 212
 QY 203 -VLVNTFDALPDALTADRY-ELIG-----IGLIPSAFLDGEDPSETSYGDLFEKSE 255
 DB 213 GVIFNSFYELES- --YVEHYTKVVGKKNWAIGPL-SLCNRDIEKAEGRKSSI-----D 264
 QY 256 ENNCVWLNKSPKSSVYVSGSVLRFPKQAEMEEIGKGLLACGRPFLMIREKND---D 312
 DB 265 EHACLKWLDSKSSIVYVCGSTADFTTAQMOELAMGLEASGQDFIWIIRTGNEWLPE 324
 QY 313 GREEEBEELSCIGELKMGKIV-SNCSOLEVLAHPALGCFVTHCGWNSAVESLSCGIPV 371
 DB 325 GFEERTKEK-----GLIIRGWAPQVILDHEAIGAFVTHCGWNSLLEGISAGVPM 374
 QY 372 VAVPQWFDQTNNAKLIEDAMGTGVRY-----RMNEGGVYDGCIEIRCVEMVMDGGDKTKL 426
 DB 375 VTPVPFAOEFNEKRLVTEVMRSAGVSGKQWKRTASEGVKREAIKAKRYM-ASEETEG 433
 QY 427 VRENAIKWKTARQAM 442
 DB 434 FRSRAKEYKEMAREAI 449

RESULT 11

W64569
 ID W64569 standard; Protein; 488 AA.
 AC W64569;
 DT 07-DEC-1998 (first entry)

DE Potato solanidine UDP-glucose glucosyltransferase.
KW Solanidine UDP-glucose glucosyltransferase; SGT; potato; tomato;
KW Glycoalkaloid; solasodine; solanaceous plant; transgenic plant.
OS Solanum tuberosum cv. Lemhi Russett.
FH Key Location/Qualifiers
FT Region 12..46
FT /note- "conserved domain"
FT Region 110..143
FT /note- "putative steroid binding region"
FT Region 351..401
FT /note- "UDP-glucose binding region"
FN W09834471-A1.
PD 13-AUG-1998.
PR 30-JAN-1998; U01864.
PR 07-FEB-1997; US-797226.
PA (USDA) US SEC OF AGRIC.
PI Allen PV, Belknap W, Friedman M, Moehs CP, Rockhold DR,
PI Stapleton A;
DR WPI; 98-446828/38.
DR N-PSDB; V49609.
DR New DNA encoding solanidine UDP-glucose glucosyl-transferase and
PT related vectors - and transgenic plants, used to reduce contents of
PT steroidal glyco-alkaloid(s) in plants, specifically potato
PS Example 1; Fig 2; 54pp; English.
CC This is the amino acid sequence of potato cv. Lemhi Russett
CC solanidine UDP-glucose glucosyltransferase (SGT). It was deduced
CC from an SGT cDNA sequence (see V49609). SGT is involved in the
CC biosynthesis of steroidal glycoalkaloids in solanaceous plants,
CC catalysing the UDP-glucose dependent conversion of the aglycone
CC solanidine to gamma-chaconine. The invention relates to DNA
CC sequences which encode SGT, and their use, particularly use of an
CC antisense DNA construct to inhibit SGT activity and glycoalkaloid
CC levels in solanaceous plants. Transgenic plants are claimed,
CC particularly tomato and potato, that have reduced contents of toxic
CC steroidal glycoalkaloids. The inactivation of glycoalkaloid
CC biosynthetic pathways is beneficial to reduce or eliminate
CC glycoalkaloid biosynthesis during storage and shipping.
CC Sequence 488 AA;
SO

Query Match 15.7%; Score 368; DB 1; Length 488;
Best Local Similarity 23.1%; Pred. No. 1.9e-29;
Matches 117; Conservative 93; Mismatches 172; Indels 124; Gaps 20;

Qy 6 VLLATPAQGHINPALQFARLLKAGDTVTFTTSVYAWRMANT----- 49
Dy 13 VLFLPLSAGHIFPLVNAARLFASRGVKATILTPHNALLFRSTIDDDVRISGFPISIVT 72
Qy 50 ---ASAAAGNPGLDFVAFSDGVDGLKPGDGKRY--MSEMKARGSEALRNLLNDDV 104
Dy 73 IKPFAEVLPEGI-----EFNSATSEMPHKIFAYLSILQRPMDKRELRPD---- 122
Qy 105 TFVYSHL-FAWAEVARLSHVPTALLWVEPATVLCIYHFYNGYADEIDAGSNEIOLP- 152
Dy 123 --CIFSDMYFPFTVDIADELHPIRLYNLSATMCYSIMH-----NLKYYRPH 167
Qy 163 RLPSEORSPTFLPATERFLMKKELETL-----DGEKAKVL 204
Dy 168 KQNLDESQ--SFVPGLEPDEIKFLSQTDLRLKSDQKTVFDELLEQVDESESYGI 225
Qy 205 V-WTFDALEPALTADRYELI-----GIGPLPSAFLEDGEDPSETSYGDDLFEP-- 254
Dy 226 VHDPTFYELEP---AYVDYTKLKKPKCMHFGL-----SHFASKIRSKELIS 269
Qy 255 EENN---CVELNLSKPKSSVYVSGSVLRFPKQMEBIGKGLLACGRPFLMWIREQKND 311
Dy 270 EHNNEIVIDLNAQPKSVLYVSGSMARFPESQLEIAQALDASNVPIFVLR----- 324
Qy 312 DGEEREELSCIGELKKMGK-----IVSWCSQLEVLAAHPALGCFVTHCGWNSAVESLSC 367
Dy 325 ---PNEETASWLPVGNLEDKTKGLYIKGWVPQLTMEHSATGFMTHCGTNSVLEAITF 381
Qy 368 GIPVAVPQWEDQTTNAKLE-----DAWCTGVRVRMNEGGGVDGCEIERCIVE-- 415

Db 382 GVPMTWPLVADQFYNEKVVGVRLGIGIKIGIDVWNEGIEI---TGPVIESAKIREAERL 438
Qy 416 MVMDDGGDKTKLVRENAIKWKTLARQA 441
Dy 439 MISNGSEELIINIRDRVMAKMAQNA 464
RESULT 12
R49245
ID R49245 standard; Protein; 471 AA.
AC R49245;
DT 08-AUG-1994 (first entry)
DE Anthocyanidin-3-glucoside rhamnosyltransferase (3RT).
KW Anthocyanidin-3-glucoside rhamnosyltransferase;
KW glucosyltransferase; inflorescence; flowering plants;
KW transgenic plant; Petunia hybrida.
OS Petunia hybrida.
FN W09403591-A.
PD 17-FEB-1994.
PR 30-JUL-1993; AU0387.
PR 30-JUL-1992; AU-003846.
PA (ITFL-) INT FLOWER DEV PTY LTD.
PI Brugliera F, Holton TA;
DR WPI; 94-065680/08.
DR N-PSDB; Q56241.
PT Nucleic acid encoding glucosyltransferase enzymes - used for
PT producing transgenic plants with altered inflorescence properties
PT including modified petal colours
PS Claim 24; Page 52-55; 76pp; English.
CC The DNA encoding the anthocyanidin-3-glucoside rhamnosyltransferase
CC is used in the construction of a vector which can then used
CC to transform plants. The transgenic plants are then capable of
CC expressing the anthocyanidin-3-glucoside rhamnosyltransferase which
CC confers altered inflorescence properties including modified petal
CC colours.
CC Sequence 471 AA;
SO

Query Match 14.1%; Score 328.5; DB 1; Length 471;
Best Local Similarity 26.6%; Pred. No. 2.2e-25;
Matches 124; Conservative 70; Mismatches 200; Indels 73; Gaps 16;

Qy 6 VLLATPAQGHINPALQFARLLKAGDTVTFTTSVYAWRMANTASAAAGNPGLDFVAF 65
Dy 12 VVMPFFFAFGHISPFVQLANKLSSYGVKVSFTTASGNASRVKSMLSA---PTTHIVPL 67
Qy 66 SDGVDDGLKPGDGKRYMSEMKARGSEALR-----NLLNDDVTFVWYSHLFAW 115
Dy 68 TLPHVEGLPGAES---TAEITPASAEKLLKVALDLMQPIKTLTLLSHLPFVDFDFAEW 124
Qy 116 AAEVARLSHVPTALLWVEPATVLCIYHFYNGYADEIDAGSNEIOLPRLPSLEORSPTF 175
Dy 125 LPKMANGLGKITYYSVVALSTAF-----LTCPARVLEPKKYPKPSLEDMKKPPL 173
Qy 176 LLPAT-----PERFLMK-----EKLET-LDGEKAKVLNVTFDALPDAL 216
Dy 174 GFPQTSVTSVTFEARDFLVFKSFHNGFTLYDRIQSLRG--CSAILAKTCSQMEGYI 231
Qy 217 TAID---RYELIGPLIPSAFLDGEDPSETSYGDDLPSEKSENNVCWMLSKPKSSVY 273
Dy 232 KYVEAQFNKPFVLIGPVVP-----DPP-----SGKLEK-----WATWLNKFFGGTVY 275
Qy 274 VSGSVLRFPKQMEBIGKGLLACGRPFLMWIREQKN-DDGEEEREEELSCIGELKKMG 332
Dy 276 CSFGSETFLTDQVKELALGLEQTGLPFELVLPANVDVSAELNRLALPEGLERVKDG 335
Qy 333 KIVS-WCSQLEVLAAHPALGCFVTHCGWNSAVESLSCGIPVAVPQWEDQTTNAKLEIDAW 391
Dy 336 IHSGWVQOOHILAHSSVGYCHAGFSSVIEALVNDQCVNMLPKQGIILNAKLVSQDM 395
Qy 392 GTGVRVRMNEGGGVDGCE-LERCVEVMVMDGDKT--KLVRENAIAKWK 435

THIS PAGE BLANK (USFTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:12 ; Search time 3616.02 Seconds
(without alignments)
1797.580 Million cell updates

Title: US-09-147-955-3
Perfect score: 1474

Sequence: 1 accaaccacaaattt.....acttaaaaaaaaaa 1474

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 104955684

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:★

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: em_gss1:*
45: em_gss2:*
46: em_gss3:*
47: em_gss4:*
48: gb_gss5:*
49: gb_gss6:*
50: gb_gss7:*
51: gb_gss8:*
52: gb_gss9:*
53: em_gss5:*
54: em_gss6:*
55: em_gss7:*
56: em_gss8:*
57: em_gss9:*
58: em_gss10:*
59: em_gss11:*
60: gb_gss10:*
61: gb_gss11:*
62: em_gss12:*
63: gb_gss12:*
64: em_gss13:*
65: em_gss14:*
66: em_gss15:*
67: em_gss16:*
68: em_gss17:*
69: em_gss18:*
70: em_gss19:*
71: em_gss20:*
72: em_gss21:*
73: em_gss22:*
74: em_gss23:*
75: em_gss24:*
76: em_gss25:*
77: em_gss26:*
78: em_gss27:*
79: gb_gss28:*
80: gb_gss29:*
81: gb_gss30:*
82: em_gss31:*
83: em_gss32:*
84: em_gss33:*
85: gb_gss34:*
86: gb_gss35:*
87: gb_gss36:*
88: gb_gss37:*
89: gb_gss38:*
90: gb_gss39:*
91: gb_gss40:*
92: gb_gss41:*
93: gb_gss42:*
94: gb_gss43:*
95: gb_gss44:*
96: gb_gss45:*
97: em_gss46:*
98: em_gss47:*
99: em_gss48:*
100: em_gss49:*
101: gb_gss50:*
102: gb_gss51:*
103: gb_gss52:*
104: gb_gss53:*
105: gb_gss54:*
106: em_gss55:*
107: em_gss56:*
108: em_gss57:*
109: em_gss58:*
110: em_gss59:*
111: em_gss60:*
112: em_gss61:*
113: gb_gss62:*
114: gb_gss63:*
115: em_gss64:*
116: gb_gss65:*
117: em_gss66:*
118: em_gss67:*
119: em_gss68:*
120: em_gss69:*
121: em_gss70:*
122: em_gss71:*
123: em_gss72:*
124: em_gss73:*
125: em_gss74:*
126: em_gss75:*
127: em_gss76:*
128: em_gss77:*
129: em_gss78:*
130: em_gss79:*
131: em_gss80:*
132: em_gss81:*
133: em_gss82:*
134: em_gss83:*
135: em_gss84:*
136: em_gss85:*
137: em_gss86:*
138: em_gss87:*
139: em_gss88:*
140: em_gss89:*
141: em_gss90:*
142: em_gss91:*
143: em_gss92:*
144: em_gss93:*
145: em_gss94:*
146: em_gss95:*
147: em_gss96:*
148: em_gss97:*
149: em_gss98:*
150: em_gss99:*
151: em_gss100:*
152: em_gss101:*
153: em_gss102:*
154: em_gss103:*
155: em_gss104:*
156: em_gss105:*
157: em_gss106:*
158: em_gss107:*
159: em_gss108:*
160: em_gss109:*
161: em_gss110:*
162: em_gss111:*
163: em_gss112:*
164: em_gss113:*
165: em_gss114:*
166: em_gss115:*
167: em_gss116:*
168: em_gss117:*
169: em_gss118:*
170: em_gss119:*
171: em_gss120:*
172: em_gss121:*
173: em_gss122:*
174: em_gss123:*
175: em_gss124:*
176: em_gss125:*
177: em_gss126:*
178: em_gss127:*
179: em_gss128:*
180: em_gss129:*
181: em_gss130:*
182: em_gss131:*
183: em_gss132:*
184: em_gss133:*
185: em_gss134:*
186: em_gss135:*
187: em_gss136:*
188: em_gss137:*
189: em_gss138:*
190: em_gss139:*
191: em_gss140:*
192: em_gss141:*
193: em_gss142:*
194: em_gss143:*
195: em_gss144:*
196: em_gss145:*
197: em_gss146:*
198: em_gss147:*
199: em_gss148:*
200: em_gss149:*
201: em_gss150:*
202: em_gss151:*
203: em_gss152:*
204: em_gss153:*
205: em_gss154:*
206: em_gss155:*
207: em_gss156:*
208: em_gss157:*
209: em_gss158:*
210: em_gss159:*
211: em_gss160:*
212: em_gss161:*
213: em_gss162:*
214: em_gss163:*
215: em_gss164:*
216: em_gss165:*
217: em_gss166:*
218: em_gss167:*
219: em_gss168:*
220: em_gss169:*
221: em_gss170:*
222: em_gss171:*
223: em_gss172:*
224: em_gss173:*
225: em_gss174:*
226: em_gss175:*
227: em_gss176:*
228: em_gss177:*
229: em_gss178:*
230: em_gss179:*
231: em_gss180:*
232: em_gss181:*
233: em_gss182:*
234: em_gss183:*
235: em_gss184:*
236: em_gss185:*
237: em_gss186:*
238: em_gss187:*
239: em_gss188:*
240: em_gss189:*
241: em_gss190:*
242: em_gss191:*
243: em_gss192:*
244: em_gss193:*
245: em_gss194:*
246: em_gss195:*
247: em_gss196:*
248: em_gss197:*
249: em_gss198:*
250: em_gss199:*
251: em_gss200:*
252: em_gss201:*
253: em_gss202:*
254: em_gss203:*
255: em_gss204:*
256: em_gss205:*
257: em_gss206:*
258: em_gss207:*
259: em_gss208:*
260: em_gss209:*
261: em_gss210:*
262: em_gss211:*
263: em_gss212:*
264: em_gss213:*
265: em_gss214:*
266: em_gss215:*
267: em_gss216:*
268: em_gss217:*
269: em_gss218:*
270: em_gss219:*
271: em_gss220:*
272: em_gss221:*
273: em_gss222:*
274: em_gss223:*
275: em_gss224:*
276: em_gss225:*
277: em_gss226:*
278: em_gss227:*
279: em_gss228:*
280: em_gss229:*
281: em_gss230:*
282: em_gss231:*
283: em_gss232:*
284: em_gss233:*
285: em_gss234:*
286: em_gss235:*
287: em_gss236:*
288: em_gss237:*
289: em_gss238:*
290: em_gss239:*
291: em_gss240:*
292: em_gss241:*
293: em_gss242:*
294: em_gss243:*
295: em_gss244:*
296: em_gss245:*
297: em_gss246:*
298: em_gss247:*
299: em_gss248:*
300: em_gss249:*
301: em_gss250:*
302: em_gss251:*
303: em_gss252:*
304: em_gss253:*
305: em_gss254:*
306: em_gss255:*
307: em_gss256:*
308: em_gss257:*
309: em_gss258:*
310: em_gss259:*
311: em_gss260:*
312: em_gss261:*
313: em_gss262:*
314: em_gss263:*
315: em_gss264:*
316: em_gss265:*
317

QY	978	aagaagaagaagtgtagttgcattggggaattgaaaaaaatggggaataagtgctg	1037
Db	343	AAGAAGACAGAAATTAAGTTGTCATGATGGAATTAAGAGAACGAAGGAAATAGTACCAT	402
QY	1038	ggtgctgcagtgtaggttcgagccacctcggttggatggtttcgtgacgcatctg	1097
Db	403	GGTGTTCAACACTTGAAGTCTCGACACATCCATCTTTAAGATGTTTGTCTCGCATCTG	462
QY	1098	ggtggaactcggctgt	1113
Db	463	GATGGAATTCGACTCT	478
RESULT	6		
AW349414/C			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

AW349414 747 bp mRNA EST 01-FEB-2000
 GM210007A20D2R Gm-rl021 Glycine max cDNA 3', mRNA sequence.
 AW349414
 AW349414.1 GI:5847124
 EST.
 soybean.
 Glycine max
 soybean.

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta: eudicotyledons: Rosidae; eurosids I; Fabales;
 Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 747)
 Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
 Erpelidg,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)
 On Jul 9, 1999 this sequence version replaced gi:5434916.
 Soybean (NSF 9872565)
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

Location/Qualifiers
 1..747
 /organism="Glycine max"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone_lib="Gm-rl021"
 /tissue_type="root"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II Xf; Site_1: EcoRI; Site_2:
 XhoI; Library Gm-rl021 is a sequence-driven, reracloned set
 of the original library Gm-cl004 which was prepared from
 root cDNA. The mRNA was isolated from entire roots of 8
 day old 'Williams' seedlings which were propagated on
 paper towels with distilled water. Stratagene's cDNA
 Synthesis Kit (catalog #200401) was used to synthesize
 the cDNA. The Gm-cl004 library was constructed by Dr.
 Paul Keim & Virginia H. Coryell, Department of Biology,
 Box5640, Northern Arizona University, Flagstaff, AZ
 86011, email: paul.keim@uau.edu, virginia.coryell@uau.edu.
 The contig analysis to select unique genes was performed
 by the laboratory of Ernest Retzel, Computational Biology
 Centers, University of Minnesota,
 http://www.cbcc.umn.edu/ResearchProjects/Soybean/index.html
 . Reracloning was performed by Genome Systems, St. Louis,
 http://www.genomesystems.com, and sequencing by the Keck
 Center for Comparative and Functional Genomics,
 University of Illinois,

FEATURES
 source

<http://www.life.uiuc.edu/biotech/keck.html>.

```

BASE COUNT      206 a 235 c 97 g 187 t 22 others
ORIGIN

Query Match      14.1%; Score 207.2; DB 71; Length 747;
Best Local Similarity 58.7%; Pred. No. 7.7e-46;
Matches 419; Conservative 0; Mismatches 253; Indels 42; Gaps 3;

QY 719 attccctccgctctcttgagcggcgaagatccctccgaaacgtctctcagcggcgatctt 778
DB 733 ATTCNNNNNGCNTTCNNNGNCGTAAGATCCCGGATACNNNTATGCGGCGTGATNNN 674
QY 779 ttcgaataatcgggaggaataactcgtggagtggtgaactcgaacggaatacttcg 838
DB 673 NNCNNTGCTC-----AATGATTAGGTGAATGCTGGACTCACGCTGAGTTATCT 620
QY 839 gtgggtgatgtgtgttgaggcgttttgaagtttccaaagccgaataatgaagatt 898
DB 619 GTGGTATTATGTTTCATTGTTACCTTGTGCTGTGTGGCTGATAGACATGAAGAGCTT 560
QY 899 gggaaaggctattagcctcgtcggaagccctttttatgatgatcagcagacagaagaat 958
DB 559 GCACGCGCGCTGCTAGATTCCGGATATCTCTCTCTGCTGGTCAATTAGAGATAG- 504
QY 959 gacgacggcgaagaagaagaagaagatgtgatgtcattgggggaattgaaaaaa 1018
DB 505 -----CAAGGAATAGAGATAAATCTGCAGAGAGGAAGAACTGGAGCAG 467
QY 1019 atggggaaaaatgtgtgtgtgtcgtcgagtgaggttctgcgcacccctgcgttgga 1078
DB 466 AGGGTAAGATTGTGAATGGTGTCTCAGGTGGAGGTCTCTCGCATGCTTCCTTGGGT 407
QY 1079 tgttctgtgaacatttggttggaactcgtgtgagaagcttgagttgaggatctcg 1138
DB 406 TCTTTTGAACGCATTGTGATGGAATTCACATATGGAAGATTGGTTCGGGGGTCTCT 347
QY 1139 gtgggtggcgtgcgcagtggttgcacgacgaagaatgcgaagctgattgagatgcg 1198
DB 346 ATGGTGGCGTTTCGCCAGTGCACAGACCAAGGACCAATGCAAGATGCTGCAAGATGT 287
QY 1199 tgggggacaggggtgagatgagaatgaa-----tgaagggggtgggttgatga 1249
DB 286 TGGAAAGACGGGGTACGGGTGGATGATGATGATGATGATGATGATGATGATGATGATG 227
QY 1250 tgtgatagaaaagtggtgtggagatgggtgatggatgggggtgacaagaccacaaactagt 1309
DB 226 GAGGAGATTAGGAAGTGTGGATGTGTGATGGGAGTGGAGGAAAGGACAGGAATC 167
QY 1310 agagaaatgccatcaaatggaagacttttgccagacagacagcatggatagatgatct 1369
DB 166 AGAAGGAATGCTGATTAATGAAATGCTTGCCCGGAGGAGCCGTCACGGAAGGAGCTCT 107
QY 1370 tcaactcaacaatccaacgcctttcttcgtcaagttgcacacttttaactctgt 1423
DB 106 TCGGATAGTAATAGGACTTTTCTCCATGATGTTGCCAAATTTGGACATGAT 53

RESULT 7
LOCUS      A1487571
DEFINITION EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            CLEDD10D21, mRNA sequence.
ACCESSION A1487571
VERSION   A1487571.1 GI:4382942
KEYWORDS EST.
SOURCE    Tomato.
ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
            Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 520)
AUTHORS   Alcalá, J., Vrebalov, J., White, R., Matern, A. L., Vision, T.,


```

```

Holt, I. E., Liang, F., Upton, J., Ronning, C. M., Craven, M. B.,
Fuji, C. Y., Bowman, C. L., Nierman, W., Fraser, C. M., Venter, J. C.,
Martin, G. B., Tankley, S. D. and Giovannoni, J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Apr 7, 1998 this sequence version replaced g1:3035667.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfri@clermson.edu
Location/Qualifiers
1..520
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED: Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 194 a 76 c 107 g 143 t
ORIGIN

Query Match      13.6%; Score 201.2; DB 40; Length 520;
Best Local Similarity 67.4%; Pred. No. 3.1e-44;
Matches 322; Conservative 0; Mismatches 138; Indels 18; Gaps 2;

QY 591 tgaaggagaagctgaaacttttagacggtgaagaaagcggaagattggtggaacagt 650
DB 57 TCAAGAGCAACTAGACACATAGATGGTGAAGAAATCAAAAGGTACTTGTGAATACAT 116
QY 651 ttgagcttgagccgcatgactcagctcttagtagtagtagtagtagtagtagtagtag 710
DB 117 TGTATGCAATAGAGCTAGAGCCACTTAACACTATTGAAATACAAATTTAATGGAATTG 176
QY 711 ggccttgattccctccgcctcttggagcggcgaagatccctccgaaacgtcttcagcgcg 770
DB 177 GACCATTTATCTCTTCATCTCTTGGTGGAAAGATTCATTGGAATCTTCATTTGGTG 236
QY 771 qcgactcttcgaaaaatcggaggaataactcgcgtgagtggtggaactcgaagccga 830
DB 237 GTGATCTTTTTCAAAAGTCAATG---ATGACATACATGGAATGTTAAACACAAAGCCTA 293
QY 831 aatctcgtggtgatgtgctgttgggagcgttttggaggtttccaaaggcacaataagg 890
DB 294 AATCATCAATTTGTTTATATCTCATTTGGGAGTCTATTGAATTTATCAAGAAACCAAAAGG 353
QY 891 aagattgggaaggcctattagcctcgggaaggccctttttagtagtagtagtagtagtag 950
DB 354 AGGAGATTGCAAAAGGGTGTATAGATGATCCAAAGGCCATCTTTATGGGTGAATAGAGATC 413
QY 951 agaagaatgacgacggcgaagagaagaagaagaagattgtagttgagttcattggggaa 1010
DB 414 A-----AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
QY 1011 tgaataaatgggaataatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1068
DB 459 TAGAGAGCAAGGGGAAATAGTACCATTGGTGTTCACAACTTGAAGTCTTGACACATCC 516

RESULT 8
LOCUS      AW442098
DEFINITION EST311494 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA

```

```

clone cLEN21G12 5', mRNA sequence.
AW442098
VERSION AW442098.1 GI:6977349
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 649)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
On Jul 7, 1999 this sequence version replaced gi:5407380.
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
Location/Qualifiers
1..649
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEN21G12"
/clone_lib="tomato fruit red ripe, TAMU"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopene accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe), 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 217 a 118 c 124 g 190 t
ORIGIN
Query Match 13.4%; Score 197.6; DB 72; Length 649;
Best Local Similarity 60.9%; Pred. No. 3.1e-43;
Matches 392; Conservative 0; Mismatches 219; Indels 33; Gaps 3;

QY 338 gtcacttgctgtactccacacctcttgcatggcgagggtggcggttgctcc 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 gtcagctccctgctgtacactctctgtctgtctgtggcagctgagcggtgaactc 60
QY 398 cagctgcagccgctctctggtgtgagccgcaccctgctgtgcataaccattc 457
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CATATCCCATCCGCGTTACTATGATGATCAACGACCACTGTGTAGACATATACTACTAT 120
QY 458 tacttaacgggtacgcagacagatgacgcgggttcccaatga-----a 502
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TACTTCAATGGCTATGAGATGAATGAAGTGTAGTACGTCAAATGATCAAAATGGAGT 180
QY 503 attcagctccctgccttccactccctgagcagcgagcttcgcagcgtttctgctgc 562
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ATCCATATGCCAAGCGCTTCCATTTACTTAAAGCCCAAGATCTCCATCTCTTTTAGTTTCA 240
QY 563 ggcagcccgagagattccggttgatgatg-----aaggaagatcgga 607
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TCTAGCTCAAAAGATGATAAGTATAGTTTCTCTACCAACATTCAAAGAGCAACTAGAC 300
QY 608 acttagacggtggaagaagcgcraagattgtgtaacacggtttgatcggttgagccc 667
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 301 ACATTAGATGGTGAAGAAAATCCAAAGGACTTGTGTAATACATTTGATGATAGACCTA 360
QY 668 gatgcactcacggctattgataggatgatcggttcggttcggttcggttcggttcggttc 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GAGCCACTCAAAAGCTATTGAAAATACAAATTAATTTGAATTTGACCATGATTCCTTCA 420
QY 728 gccttttggacggcgaagatccctccgaacgctcttacggcgccgctcttttcgaaaaa 787
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TCATTTCTGGGTGGGAAAAGATTCAATTTGAATTTCTTCAATTTGGTGGTATCTTTTCAAAAG 480
QY 788 tcggagagagaataactcgtcgtgagtggttgaaactcgaagccgaatacttcggttggtat 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TCAATG---ATGACTACATGGAATGGTTAAACACAAAGCCTAATCATCAATTTGTTAT 537
QY 848 gtctcgtttggagcgttttttgagtttccaaagccaaatggaagatggaagatggaaggg 907
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 ATCTCAATTTGGGAGTCTATGAAATTTATCAAGAACCAAGAGGAGGATTCGAAAGGG 597
QY 908 ctattagcctgcggaaggccctttttatggatgatacagaaca 951
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 TTGATAGATCCAAAAGCCATCTTATGGTAAATAAGATCA 641

RESULT 9
LOCUS AI489014 500 bp mRNA EST 29-JUN-1999
DEFINITION EST247353 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
CLEDI9F7, mRNA sequence.
ACCESSION AI489014
VERSION AI489014.1 GI:4384385
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 500)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Nov 2, 1998 this sequence version replaced gi:3830982.
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.
Location/Qualifiers
1..500
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEDI9F7"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 188 a 69 c 105 g 138 t
ORIGIN
Query Match 13.1%; Score 193.2; DB 40; Length 500;
Best Local Similarity 67.3%; Pred. No. 4.6e-42;
Matches 311; Conservative 0; Mismatches 133; Indels 18; Gaps 2;

```


Page 9

```

EST259978 tomato susceptible, Cornell Lycopersicon esculentum CDNA
clone CLES719, mRNA sequence.
ACCESSION
  AI779099
  AI779099.1 GI:5277140
  EST.
SOURCE
  tomato.
ORGANISM
  Lycopersicon esculentum
  Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
  Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
  Solanaceae; Solanum; Lycopersicon.
  1 (bases 1 to 517)
  D' Ascenzo, M., He, X., Lyman, J., Matern, A. L., Vision, T., Holt, I. E.,
  Liang, F., Upton, J., Ronning, C. M., Craven, M. B., Fujii, C. Y.,
  Bowman, C. L., Nierman, W., Fraser, C. M., Venter, J. C., Tankley, S. D.,
  Giovannoni, J. J. and Martin, G. B.
  Generation of ESTs from Pseudomonas susceptible tomato
  Unpublished (1999)
  On Jan 19, 1998 this sequence version replaced gi|2285531.
  Contact: David Frisch
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 4366
  Fax: 864 656 4293
  Email: dfrisch@CLEMSON.EDU
  5 prime sequence.
  Location/Qualifiers
    1..517
    /organism="Lycopersicon esculentum"
    /cultivar="RII-13 (Rio Grande x Money Maker)"
    /db_xref="taxon:4081"
    /clone="CLES719"
    /clone_lib="tomato susceptible, Cornell"
FEATURES
  source

```

```

/clone_illd="tomato susceptible", Cornell
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/note="vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLES - Tomato Pseudomonas Susceptible EST Library.
Directionally cloned cDNAs inserted into pBlueScript
SK(-) at 5' end with EcoRI and 3' end with XhoI site"
BASE COUNT      185 a   78 c   102 g   152 t
ORIGIN

Query Match          11.6%; Score 170.6; DB 44; Length 517;
Best Local Similarity 69.0%; Pred. No. 6.5e-36;
Matches 249; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY    591 tgaagagaagctggaactttagacgggtgaagagaagcgcaagtattcggtagaacagt 650
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    152 TCAAGAGCAACTAGACACATAGATGCTGAAGAAATCCAAAGGTACTTGTGAATACAT 211

QY    651 ttgatcgttgagccgcgatcacctcagcgctatgataggtagttagtcgggatacgcg 710
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    212 TTGATGCATTAGCGTAGGCCATCAAAGCTATTGNAATAACAATTTTAATGGAAATG 271

QY    711 ggccgttgattcccctccctcttgtagcgccaagatccccctcgaaacctctaccgag 770
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    272 GACCATTGATCCTTCATCATCTTGCTGGTGAAAAAAGATTCAATTTGGAAATCTTCATTTGGTGTG 331

QY    771 gcgatctttcgaaaaatcgagagagataactcgtggagtggttgaactcgaaagccga 830
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    332 GTGATCTTTTTCAAAGGTCAAATG---ATGACTACATGGGAATGGTTAAACACAAAAGCCCTA 388

QY    831 aaatccggtgggtgatgctgttgttgggagcgcttttgaggtttccaaaggcacaatacg 890
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    389 AATCATCAATTGTTTATATCATTCATTTGGGAGCTCTATTGAATTTATCAAGAAACCAAAAGG 448

QY    891 aaagatatgggaaaaagggtatagccctgcggagaagccccttttatggatgatcacgagac 950
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    449 AGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATTTCTTTATGGGTAATAAGAGATC 508

QY    951 a 951

```

Db 449 AGGAGATTGCAAAAGGTTGATAGATCCAAAGGCCATTCTTATGGTAATAAGAGATC 508

Qv 951 a 951

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:31 ; Search time 61.65 Seconds
(without alignments)
517.335 Million cell updates/sec

Title: US-09-147-955-2
Perfect score: 2421
Sequence: 1 MVRVRLATFAQGHINDA.....GEGSSKLNAFLRQVARA 450

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL12:*
- 1: sp.archaea:*
 - 2: sp.bacteria:*
 - 3: sp.fungi:*
 - 4: sp.human:*
 - 5: sp.invertebrate:*
 - 6: sp.mammal:*
 - 7: sp.mhc:*
 - 8: sp.organelle:*
 - 9: sp.phage:*
 - 10: sp.plant:*
 - 11: sp.todent:*
 - 12: sp.virus:*
 - 13: sp.vertebrate:*
 - 14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2421	100.0	460	10	Q9r27 perilla fru
2	2223	91.8	443	10	Q9r26 perilla fru
3	1646.5	68.0	461	10	Q9r25 verbena x h
4	1053.5	43.5	467	10	P93709 nicotiana t
5	1038	42.9	456	10	Q23270 arabidopsis
6	995	41.1	519	10	Q9zy2 arabidopsis
7	971	40.1	474	10	O04930 arabidopsis
8	949.5	39.2	455	10	Q9zy5 arabidopsis
9	888	36.7	458	10	Q23406 arabidopsis
10	716.5	29.6	449	10	O22822 arabidopsis
11	702	29.0	479	10	O23401 arabidopsis
12	687.5	28.4	449	10	O22820 arabidopsis
13	675.5	27.9	460	10	O48676 arabidopsis
14	663.5	27.4	456	10	O22182 arabidopsis
15	658.5	27.2	484	10	O23400 arabidopsis
16	653.5	27.0	438	10	O22183 arabidopsis
17	628.5	26.0	475	10	O23402 arabidopsis
18	604.5	25.0	453	10	O22186 arabidopsis
19	580.5	24.0	481	10	Q9zwj3 arabidopsis

20	484	20.0	455	10	O64733 arabidopsis
21	481.5	19.9	482	10	Q9ZUV0 arabidopsis
22	455.5	18.8	440	10	O64732 arabidopsis
23	446.5	18.4	477	10	Q9ZWK5 vigna mungo
24	431.5	17.8	491	10	Q9ZQ99 arabidopsis
25	430	17.8	454	10	Q9XF16 forsythia x
26	428	17.7	496	10	Q9ZQ97 arabidopsis
27	421	17.4	484	10	Q9ZQ94 arabidopsis
28	415	17.1	496	10	Q9ZQ96 arabidopsis
29	414.5	17.1	462	10	O04622 arabidopsis
30	413	17.1	347	10	O49042 pisum sativ
31	409.5	16.9	455	10	Q9ZWS2 vigna mungo
32	409	16.9	495	10	Q9ZQ95 arabidopsis
33	408.5	16.9	447	10	O04114 perilla fru
34	407.5	16.8	452	10	O48715 arabidopsis
35	406.5	16.8	466	10	O43526 lycopersico
36	404.5	16.7	495	10	Q9ZQ94 arabidopsis
37	400.5	16.5	476	10	P93364 nicotiana t
38	398.5	16.5	420	10	O22603 ipomoea pur
39	398	16.4	452	10	O23380 arabidopsis
40	397	16.4	496	10	Q9ZQ98 arabidopsis
41	396.5	16.4	478	10	O49492 arabidopsis
42	394.5	16.3	280	10	Q9ZWK4 vigna mungo
43	392.5	16.2	476	10	P93365 nicotiana t
44	391	16.2	488	10	P93789 solanum tub
45	387	16.0	474	10	O82382 arabidopsis

ALIGNMENTS

RESULT	1				
ID	Q9Zr27	PRELIMINARY;	PRT;	460	AA.
AC	Q9Zr27				
DT	01-MAY-1999 (TREMBLrel. 10, Created)				
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)				
DE	UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.				
GN	PF3R4				
OS	Perilla frutescens.				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;				
OC	core eudicots; Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-CV, SHIKUN; TISSUE=LEAF;				
EX	MEDLINE; 99167509.				
RA	YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,				
RA	KUSUMI T., SAITO K.;				
RT	*Molecular cloning and biochemical characterization of a novel				
RT	anthocyanin 5-O-glucosyltransferase by mRNA differential display for				
RT	plant forms regarding anthocyanin.;				
RL	J. Biol. Chem. 274:7405-7411(1999).				
DR	EMBL; AB013596; BAA36421.1; -				
DR	PROSITE; PS00375; UDPGT; 1.				
KW	Transferase.				
SW	SEQUENCE 460 AA; 50974 MW; FE7CCF22 CRC32;				

Query Match 100.0%; Score 2421; DB 10; Length 460;
Best Local Similarity 100.0%; Pred. No. 6.1e-190;
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MVRVRLATFAQGHINDA.....GEGSSKLNAFLRQVARA	60
DB	1	MVRVRLATFAQGHINDA.....GEGSSKLNAFLRQVARA	60
QY	61	DFVAFSDGDDGLKPCGDKGKRYMSEKARGSEALRNLLNNHDDVTFVYVYSHLFAWAEVA	120
DB	61	DFVAFSDGDDGLKPCGDKGKRYMSEKARGSEALRNLLNNHDDVTFVYVYSHLFAWAEVA	120
QY	121	RESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLPPLPLEQRSLFTLLPET	180

```

|||||
121 RESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLRPLPLEORSPLTFLPET 180
QY 181 PERFLMMKKELETLDGEEKAKVLVNTFDALPDALTAIDRYELIGIGLPLPSAFLOGGD 240
Db 181 PERFLMMKKELETLDGEEKAKVLVNTFDALPDALTAIDRYELIGIGLPLPSAFLOGGD 240
QY 241 PSETSYGGDLFEKSEENNCVWLDTPKSSVYVSGSVLRPFPKQMBEIGKGLLACGRP 300
Db 241 PSETSYGGDLFEKSEENNCVWLDTPKSSVYVSGSVLRPFPKQMBEIGKGLLACGRP 300
QY 301 FLWMIREQKNDGEEEBEELSCIGELKMKGIKIVSWCQLEVLHAPALGCFVTHCGWNSAV 360
Db 301 FLWMIREQKNDGEEEBEELSCIGELKMKGIKIVSWCQLEVLHAPALGCFVTHCGWNSAV 360
QY 361 ESSLSCGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRNMGEGGVDSIERCVEMVMDGGE 420
Db 361 ESSLSCGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRNMGEGGVDSIERCVEMVMDGGE 420
QY 421 KSKLVRENAIKWKTLLAREAMGEDSSSLKNLNAFLHQVARA 460
Db 421 KSKLVRENAIKWKTLLAREAMGEDSSSLKNLNAFLHQVARA 460

RESULT 2
Q9ZR26
ID Q9ZR26 PRELIMINARY; PRT; 443 AA.
AC Q9ZR26;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCTANIN 5-O-GLUCOSYLTRANSFERASE.
GN PF3R6.
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiales; Perilla.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SHIKUN; TISSUE=LEAF;
RX MEDLINE: 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013597; BAA36422.1; -.
DR PROSITE; PS00375; UDFGT; 1.
KW Transferase.
SQ SEQUENCE 443 AA; 49110 MW; D87B5486 CRC32;

Query Match 91.8%; Score 2223; DB 10; Length 443;
Best Local Similarity 95.7%; Pred. No. 8.9e-174;
Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;

QY 1 MYRRVLLATPPAQGHINPALQFAKRLKAGTDVTFPTSYYAWRRMANTASAAAGNPGL 60
Db 1 MYRRVLLATPPAQGHINPALQFAKRLKAGTDVTFPTSYYAWRRMANTASAAAGNPGL 60
QY 61 DFVAFSDGYDDGLKPCGDKGRYKRYMSEMKARGSEALRNLLNNHDDTVFVYSHLFAWAAEVA 120
Db 61 DFVAFSDGYDDGLKPCGDKGRYKRYMSEMKARGSEALRNLLNNHDDTVFVYSHLFAWAAEVA 120
QY 121 RESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLRPLPLEORSPLTFLPET 180
Db 121 RLSHVPPTALLWVEPATVLCIYYFYFNGYADEIDAGSNEIQLRPLPLEORSPLTFLPAT 180
QY 181 PERFLMMKKELETLDGEEKAKVLVNTFDALPDALTAIDRYELIGIGLPLPSAFLOGGD 240
Db 181 PERFLMMKKELETLDGEEKAKVLVNTFDALPDALTAIDRYELIGIGLPLPSAFLOGGD 240

```

```

QY 241 PSETSYGGDLFEKSEENNCVWLDTPKSSVYVSGSVLRPFPKQMBEIGKGLLACGRP 300
Db 241 PSETSYGGDLFEKSEENNCVWLDTPKSSVYVSGSVLRPFPKQMBEIGKGLLACGRP 300
QY 301 FLWMIREQKNDG--EEEBEELSCIGELKMKGIKIVSWCQLEVLHAPALGCFVTHCGWNS 358
Db 301 FLWMIREQKNDGEEEBEELSCIGELKMKGIKIVSWCQLEVLHAPALGCFVTHCGWNS 360
QY 359 AVESLSCGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRNMGEGGVDSIERCVEMVMDG 418
Db 361 AVESLSCGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRNMGEGGVDSIERCVEMVMDG 420
QY 419 GEKSLVRENAIKWKTLLAREAMG 441
Db 421 GDKTKLVRENAIKWKTLLARQAMG 443

RESULT 3
Q9ZR25
ID Q9ZR25 PRELIMINARY; PRT; 461 AA.
AC Q9ZR25;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCTANIN 5-O-GLUCOSYLTRANSFERASE.
GN HGT8.
OS Verbena x hybrida.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Verbenaceae;
OC Verbena.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PETAL;
RX MEDLINE: 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013598; BAA36423.1; -.
DR PROSITE; PS00375; UDFGT; 1.
KW Transferase.
SQ SEQUENCE 461 AA; 51347 MW; 6DCBC3E6 CRC32;

Query Match 68.0%; Score 1646.5; DB 10; Length 461;
Best Local Similarity 70.0%; Pred. No. 1.3e-126;
Matches 326; Conservative 49; Mismatches 74; Indels 17; Gaps 7;

QY 1 MYRRVLLATPPAQGHINPALQFAKRLKAGTDVTFPTSYYAWRRMANTASAAAGNPGL 60
Db 1 MSRAHVLLATPPAQGHINPALQFAKRLANADIQTFTPTSYYAWRRMSRT---AAGSNGLI 57
QY 61 DFVAFSDGYDDGLKPCGDKGRYKRYMSEMKARGSEALRN--LLNNHDD---VTFVYSHLPA 114
Db 58 NFVFSFDGYDDGLQPGDDGKNYSEMKSRGKLSLSDTLAANNVDQKSKITFPVYSHLPA 117
QY 115 WAAEVARESQVPSALLWVEPATVLCIYYFYFNGYADEIDAGSDEIQLR--RLPLEORSPL 173
Db 118 WAAKVAAREFHLRSALLWTEPATVLDIFVFYFNGYDEIDAGSDAIHLPGGLPVAQROLP 177
QY 174 TFLPETPERFLMMKKELETLDGEEKAKVLVNTFDALPDALTAIDRYELIGIGLPLPS 233
Db 178 SFLLPSTHERFSLMKKELETLEGEERPKVLVNSFDALPDALKADIKTEMTAIGLPLPS 237
QY 234 AFLDGDGPSETSYGGDLFEK--SEENNCVWLDTPKSSVYVSGSVLRPFPKQMBEIGK 292
Db 238 AFLDGDKPSDRSFGDLFEKSGNDDCLEWLSTNPRSSVYVSGSVFVNTTKSQMBEIA 297
QY 293 GLLACGRPFLWMIREQKNDGEEEBEELSCIGELKMKGIKIVSWCQLEVLHAPALGCFVT 352

```

```

Db 298 GLLDCGRPLWVRV-----NEGEBVLSCHEELKRVGKIVSCSLEVLTHPSLGCFTV 352
QY 353 HCGWNSAVESLSCGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRMNEGGS-VDGSEIERC 411
Db 353 HCGWNSLSEISFGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRMNEGGSVVDGDBIRRC 412
QY 412 VEMVMDGGKSKLVRENAIKWKTILAREAMGDSGLKLNALFLHOV 457
Db 413 IEEVMDGGKSKLVRENAIKWKTILAREAMGDSGLKLNALFLHOV 458

RESULT 4
ID P93709 PRELIMINARY; PRT; 467 AA.
AC P93709;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE GLUCOSYL TRANSFERASE.
GN JIGT.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BY-2; TISSUE-SUSPENSION CULTURE;
RA KOJIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB000623; BAA19155.1; -.
DR MENDEL: 9421; Nicta.1525; 9421.
DR PROSITE: PS00375; UPGT; 1.
DR PFAM: PF00201; UPGT; 1.
KW transferase.
SQ SEQUENCE 467 AA; 52989 MW; 3726B31 CRC32;

Query Match 43.5%; Score 1053.5; DB 10; Length 467;
Best Local Similarity 45.3%; Pred. No. 3.8e-78;
Matches 209; Conservative 83; Mismatches 150; Indels 19; Gaps 7;

QY 6 VLLATPPAGCHINPALQFAKRLKAGTDVFTFTSVYAWRRMANTASAAAGNPPGLDFVAF 65
Db 10 VLLATPPAGCHINPALQFAKRLKAGTDVFTFTSVYAWRRMANTASAAAGNPPGLDFVAF 64
QY 66 SDGYDGLKPCGDGKR-VYSEMKARGSEALRNLL-----LNHVDVTFVYSHLFAWAAEVA 120
Db 65 SDGYDGNFGSDDYHLNSAISKSHGSEFFIANLKSARKNGYPTTRVITILMDWAGVA 124
QY 121 RESQVPSALLWVEPATVLCIYFYFNGVAD---BIDAGSEIQLPPLPORSPTFL 177
Db 125 KKLHIPSTLFTQPATVFDIYYFTNFNFYFNKYDSQDQIIEPLGLSLSSDPSFVF 184
QY 178 PETPERFLM--MKEKLETLDGEKAKVLVNTFDALPDALTAIDRYELIGLPIPSAF 235
Db 185 DVKSNDAVESIKQIEILNSEENPRILVNTFDALBNALRVKNTWVGTLIPSF 244
QY 236 LOGDPSSETS YGGDLFPKSENNCVENLDTPKSSVYVFGSVLRPPAKQMEIGKGL 295
Db 245 LDEKDKONFFAADMIE--SENNYEMWLDARANKSVIYAFGSYAEISSQWMEISQGL 302
QY 296 AGRPFLMIREOKNDGEEEEEELSCIGELKMGKIVSCSLEVLTHPSLGCFTVHCG 355
Db 303 KGRFLWVIRETLN--GEKPEKLTCKDELEKIGRIVRWCSEHVEKHSVCCFLTHCG 360
QY 356 WNSAVESLSCGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRMNEGGSVVDGDBIRRC 415
Db 361 WNSTLESASGVPVAVPQWFQDTTNAKLIEDAWGTGVRVRMNEGGSVVDGDBIRRC 420
QY 416 MDGGEKSKLVRENAIKWKTILAREAMGDSGLKLNALFLHO 456
Db 421 MDAEEGELRLKNAOKWDLAKESKTSKSSNNVNLKAYNE 461

```

```

RESULT 5
ID Q23270 PRELIMINARY; PRT; 456 AA.
AC Q23270;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE GLUCOSYL TRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE; 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERGKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFANELLI P., WEDLER H.,
RA WEDLER E., WAMBUTT R., WEITZENEGGER T., POHL T.M., TERRY N.,
RA GLENN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOULELATOU E., MILIONI D., HATZIOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana."
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z57335; CAB10189.1; -.
DR PROSITE: PS00375; UPGT; 1.
DR PFAM: PF00201; UPGT; 1.
KW transferase.
SQ SEQUENCE 456 AA; 50702 MW; EAE0D380 CRC32;

Query Match 42.9%; Score 1038; DB 10; Length 456;
Best Local Similarity 48.2%; Pred. No. 6.7e-77;
Matches 225; Conservative 72; Mismatches 134; Indels 36; Gaps 12;

QY 3 RRVLLATPPAGCHINPALQFAKRLKAGTDVFTFTSVYAWRRMANTASAAAGNPP---G 59
Db 11 RPHYLLVTPPAOCHINPALQANRLIHGATVTVSTAVSAHRM-----GEPSTKG 62
QY 60 LDFVAFSDGYDGLKPCGDGKR-VYSEMKARGSEALRNLLNHD-----VTFVYSHL 112
Db 63 LSPAFWTFDGDGLSKSFEQKIYMSLRCGNSALRIKANDLATETETPTGVIVSVL 122
QY 113 FAWAAEVARESQVPSALLWVEPATVLCIYFYFNGVAD---BIDAGSEIQLPPLPORS 171
Db 123 VPMVSTVAREFHLPPTLLWIEPATVLDIYYFTNFNFYFNKYDSQDQIIEPLGLSL 180
QY 172 LPTFLPETPERFLM--MKEKLETLDGEKAKVLVNTFDALPDALTAIDRYELIGL 230
Db 181 LPSFLQPSKALPSALVTRHEHTALETESNFKILVNTFSALEHDLITSVEKLMIDIGPL 240
QY 231 IPSAFLOGDPSSETS YGGDLFPKSENNCVENLDTPKSSVYVFGSVLRFPKQAMEE 289
Db 241 VSSS-----EGKT---DLFKSDE-DYTKWLDKSLERSVIYISLTGTHADDLPKHEA 289
QY 290 IGGGLACCRPFLMIREOKNDGEEEEEELSCIGELKMGKIVSCSLEVLTHPSLGC 349
Db 290 LTHGVLATNRPFLLWVIRE--KNPEKKKRNFLLELRGSDR--GLVVGWCSQVAVLAHCAVGC 347

```

```

QY 350 FVTHCGHNSAVESLCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRNMGCGVDGSGEIE 409
Db 348 FVTHCGHNSLESVGPVAVPQWFDQTTNAKLIEDAWGTGVRVRNMGCGVDGSGEIR 407

QY 410 RCVEVMMDGGESKLVRENAIKWTLAREAMGDDGSSLNKNAFLHQ 456
Db 408 RCLEKVMGSGEAEEMRENAEKWMAVDAABGGPSDLNKGFDVE 454

RESULT 6
Q92VY2 PRELIMINARY; PRT; 519 AA.
AC Q92VY2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE T25N20.20;
DE Arabidopsis thaliana (Mouse-ear cross).
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,
RA VISOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT *Genomic sequence for Arabidopsis thaliana BAC T25N20.*;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80599.1; -
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 519 AA; 58584 MW; C982007F CRC32;

```

```

Query Match 41.1%; Score 995; DB 10; Length 519;
Best Local Similarity 45.8%; Pred. No. 2.7e-73;
Matches 213; Conservative 75; Mismatches 143; Indels 34; Gaps 11;

QY 7 LIATPPAGGHINPALQFAKRLK--AGTDVTFFTSVYAMRMANTASAAAGNPPGLDFVAF 65
Db 57 LLVTFPAGGHVNPISLRFARLIRKTRGTFTVCVSPH---NSMIANHNKVENLSFLTF 113

QY 66 SDGYDD-GLKCGCGKRYMSEMKARGSEALRNLL--LNND--VTFVYVSHLFAWAAEVA 120
Db 114 SDGFDGDIETDROKRSVNLKVGNDKALSDPIEATKNGDSPVTCLITILLNAPKVA 173

QY 121 RESQVPSALLWVEPATVLCIYFFYFNGYADEIDAGSDEIQLPRLPBOERSLPTLLPET 180
Db 174 RRFQLPALLWIQPALVFNYYTHFG-----NKSVPFLNLSLEIRLPSLTPSN 226

QY 181 PERPRL-MMKEKLETLDGEEKAKVLYNTFDALPDALTADRYELIGLPLPSAFLDGG 239
Db 227 TNKGAYDAFOEMEFELIKETKPKILINTFDSLEPEALTAFPNIDMVAVGPLLPTEIFSGS 286

QY 240 DPSETS YGDLFEKSENNCVMELDKPKSSVYVSGVLRFPRAQMEIGKGLLAGCR 299
Db 287 TNKSV-----KQSSSYTLMDSKTESVIYVSGFTWVLSKKQIEELARALIEGR 338

QY 300 PFLWMIROKN-----DGEDEEELSICG---ELKMGKIVSWCSOLEVLAHPALGCVPT 352
Db 339 PFLAWITOKSNRETKEGESETETERKTAGFRHELEEVGMIVSWCSOIEVLSHRAGCVPT 398

QY 353 HCGMNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRNMGCGVDGSGIERCV 412
Db 399 HCGWSTLESILVGPVAVPQWFDQTTNAKLIEESMKTGVRVRNMGKGLVERGERICL 458

QY 413 ENVMDDGSKLVRENAIKWTLAREAMGDDGSSLNKNAFLHQV 457
Db 459 EAYME--EKSVELRENAKKRLAMEAGREGGSSDKNMEAFVEDI 501

```

RESULT 7

```

O04930 PRELIMINARY; PRT; 474 AA.
AC O04930;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
DE (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE).
GN TAGLJ.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA GILAD A., SCOLNIK P.A., BAR-2VI D.;
RL Plant Physiol. 113:1004-1004 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA THORNBERG R.W., GRAHAM R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOLE-3-ACETATE = UDP +
CC INDOLE-3-ACETYL-BETA-1-D-GLUCOSE.
DR EMBL; U01293; AB58497.1; -
DR MENDEL; 16583; Arabid; 2543; 16583.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
DR Transferase; Glycosyltransferase.
KW SEQUENCE 474 AA; 53740 MW; 35521D70 CRC32;

```

```

Query Match 40.1%; Score 971; DB 10; Length 474;
Best Local Similarity 44.9%; Pred. No. 2.1e-71;
Matches 216; Conservative 74; Mismatches 137; Indels 54; Gaps 17;

QY 7 LIATPPAGGHINPALQFAKRLK--AGTDVTFFTSVYAM--RMANTASAAAGNPPGLDFV 63
Db 15 LVTFPAGGHVNPISLRFARLIRKTRGTFTVCVSPH---NSMIANHNKVENLSFLTF 70

QY 64 AFSQGVDDGLKPC--GDKR-----YNSMKARGSEALRNLL-----LNNDHDTFVVVSH 111
Db 71 TYSDDGDDGFKSSAYSOKSDATGTFNPMSEMRRGKETTELIEDNRKQNRPFCTCVITI 130

QY 112 LFAWAAEVARESQVPSALLWVEPATVLCIYFFYFNGYADEI-----DAGSDEIQLPRLPPL 167
Db 131 LLTWVAELAREFHLPALLWQPVVPSIFVHYFNGYDAISEMANTPSSIKLPSLPL 190

QY 168 EORSPLTLLPETPERFL-MMKEKLETLDGEEKAKVLYNTFDALPDALTAL-DRYELI 225
Db 191 TVRDIPSPISVSNVAFLLPAFREQIDSLKEENPKILINTFOELEPEAMSVPPNFIV 250

QY 226 GIGLIPSAFLDGGDPSETS YGDLFEKSENNCVMELDKPKSSVYVSGVLRFPKPA 285
Db 251 PVGPLL-----TLRTDF-----SSRGEVIEWLDTKADSSVLYVSGTFLAVLSKK 294

QY 286 OMEETGKGLACGPFPLWMTREQ--KNDGGEFEELSCIGELKK---MGKIVSWCSOL 339
Db 295 QLVLELCKALISRRPFLWITDKSYRNKKEDEKEER-DCISSEKSFDEIGVWVSWCQDF 353

QY 340 EYLAHPALGCVPTHCGMNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVR--RM 397
Db 354 RVLNHRISGCVPTHCGMNSTLESILVGPVAVPQWFDQTTNAKLIEDCKWTKGVRVMEKK 413

QY 398 NEGQ--GYDGSIEIRCEVEMVNDGCKSKLVRENAIKWTLAREAMGDDGSSLNKNAFLH 455
Db 414 EREGVVVDVSEIRRCIEEYNE--DKAEERFGNATRWKDLAAEVREGGSSFNHLKAFVD 471

QY 456 Q 456
Db 472 E 472

```

```

RESULT 8
Q92VY5 PRELIMINARY; PRT; 455 AA.
AC Q92VY5
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE T25N20.17.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,
RA VYSOTSKAJA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T25N20." ;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80596.1; -
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 455 AA; 51190 MW; F9BAC33E CRC32;

Query Match 39.28; Score 949.5; DB 10; Length 455;
Best Local Similarity 44.08; Pred. No. 1.2e-69;
Matches 206; Conservative 78; Mismatches 153; Indels 31; Gaps 10;

QY 1 MYRRRLVLTTPAQGHINPQAFKRLK--AGTDVTFVTSVYARRMANTASAAAGNPPG 59
Db 1 MAQPHLLVTPAQGHVNPRLREARRLKITGARTVATCLSVHR---SMIPNNHNVEN 57
QY 60 LDFVAFSDGDDG-LKPCGDKRYMSEMKARGSEALNLLNHD----VTFVYSHLFA 114
Db 58 LSFLLTSDGDFDGVTSNDDQVNLVHFERNKGKALSDFEANQNGDSPVSCLIYTLPN 117
QY 115 WAAEVAEASQVPSALLWVEPATVLCIYFFYFNGYADIDAGSDIQLPLPLEORSIPT 174
Db 118 WPKVAFRFLPSVHLWQAPAFEDIIYNYSTG-----NNSVFEPNLSLEIRDLPS 170
QY 175 FLPL-ETPERFLMKKLETLDGEEKAKVLVNTFDALPDALTAIDRVELIGIPLPS 233
Db 171 FLSPNTNKAQVTOELMDLKEENPKILVNTFDSLEPEFTLAIPTNEMVAVGELLPA 230
QY 234 AFLDGDPSSETSYGDLPEKSEENNCVEMLDTPKSSVYVVSFGSVLRFPKQMEIEIGK 293
Db 231 EIFTGSES-----GKLSRDHQSSYTLMLDSKTESVIVYSGTWSVLSKKQIEELARA 285
QY 294 LIACGRPFLWMIREOKND----DGEDEEEELSCIG---ELKMGKIVSCSOLVLAHPA 346
Db 286 LIEGRPFLWITDKLREAKIEGEBETEIEKTAGFHELEERYGMIVSCSOLVLRHRA 345
QY 347 LACFVTHCGWNSAVESLSCGVPVYVAVPQWFDQTTNAKLIIDAWGTGVRVMEGGVDGS 406
Db 346 IGCFLTHCGWSSLESVLGVYVAVPQWSDQPANAKLLEEIKWTKGVRVRENSGLVIRG 405
QY 407 EIERCVENMDGGEKSLVRENATKWTKLAREANGEDGSSLKLNLAFL 454
Db 406 EIMRCLEAVMEA--KSVLEIRENAEKWRKLATAGREGGSSDKNVEAFV 451

```

```

RESULT 9
O23406 PRELIMINARY; PRT; 458 AA.
AC O23406;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE SIMILAR TO INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA KEVAN M., STIEKEMA W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,
RA JONES J., PALME K., ANSORGE W., DEISENY M., BANCROFT I., MEWES H.W.,
RA SCHUELLER C., CHALWATZIS N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 297339; CAB10333.1; -
DR MENDEL; 26727; Arath; 2543.26727.
DR PROSITE; PS00375; UDPGT; 1.
DR FRAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 458 AA; 52005 MW; 67943A2F CRC32;

Query Match 36.78; Score 888; DB 10; Length 458;
Best Local Similarity 42.48; Pred. No. 1.2e-64;
Matches 204; Conservative 73; Mismatches 134; Indels 70; Gaps 18;

QY 7 LIATPFAQGHINPQAFKRLK--AGTDVTFVTSVYARRMANTASAAAGNPPGLDVF 63
Db 15 LFVTPFAQGHINPQAFKRLK--AGTDVTFVTSVYARRMANTASAAAGNPPGLDVF 63
QY 64 AFSGDGDDGLKPC--GDGKR-----YMSKARGSEALNLL----LNHDTVYVYSH 111
Db 71 TVSDGHDGDFKSSAYSDKSRQDAGTGNFMSEMRRGKFTLTELIEDNRKQRPFCVYVI 130
QY 112 LFAWAAEVAESQVPSALLWVEPATVLCIYFFYFNGYADEI---DAGSDEIQLPLPL 167
Db 131 LLTWAEALA-----LFSIFVHYFNGYEDALISEMANTPSSSIKPLSLPL 174
QY 168 EORSPTFLPLPETPERFL--MKKLETLDGEEKAKVLVNTFDALPDALTAIDRVELI 225
Db 175 TVNDIPSFVIVSNVYAFLLPAPREQIDSLKEEINPKILINTFOLEPEAMSVDPNFKIV 234
QY 226 GIGPLPSAFLDGDPSSETSYGDLPEKSEENNCVEMLDTPKSSVYVVSFGSVLRFPKA 285
Db 235 PVGPLL-----TLRTDF-----SSRGEVIELDTKADSSVLYVSFGTLAVLSKK 278
QY 286 QMEIEIGKGLACGRPFLWMIREQ--KNDGDEEEELSCIG---ELKMGKIVSCSOL 339
Db 279 QLVELCKALIQSRPFLWITDKSYRNKEDQEKEE--DCISSFREELDEIGMVVSCDQF 337
QY 340 EYLAHPALGCFVTHCGWNSAVESLSCGVPVYVAVPQWFDQTTNAKLIIDAWGTGVRV--RM 397
Db 338 RVLNHRISGCFVTHCGWNSLLESVSGVPVYVAVPQWSDQPANAKLLEEIKWTKGVRVMEKK 397
QY 398 NEGQ--GVDPGSIEICVEMVMDGGEKSLVRENATKWTKLAREANGEDGSSLKLNLAFL 455
Db 398 EEEGVVVVDSSEIRRCIEVME--DKAEFRGNATRWKDLAAEAVREGGSSPHILAKAFVD 455
QY 456 Q 456
Db 456 E 456

RESULT 10
O22822 PRELIMINARY; PRT; 449 AA.
AC O22822;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE ISOLOG.
OS Arabidopsis thaliana (Mouse-ear cross).

```


Query Match 27.9%; Score 675.5; DB 10; Length 460;
Best Local Similarity 34.9%; Pred. No. 2.9e-47;
Matches 169; Conservative 72; Mismatches 186; Indels 57; Gaps 15;

RESULT	14
022182	
ID	PRELIMINARY; PRT; 456 AA.
AC	022182;
DT	01-JAN-1998 (TremBrel. 05, Created)
DT	01-JAN-1998 (TremBrel. 05, Last sequence update)
DE	01-NOV-1999 (TORE3-Last annotation update).
DE	PUTATIVE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
GN	T20D16.11.
OC	Arabidopsis thaliana (Mouse-ear cress).
OC	Arabiopsids
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC	cory eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
RN	Arabidopsis.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	STRAIN-CV. COLUMBIA;
RA	ROUNSLBY S.D., LIN X., KETCHUM K.A., CROSSBY M.L., BRANDON R.C.,
RA	SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA	SOMERVILLE C.R., VENTER J.C.;
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AC002391; AAC87119.1; -
DR	MENDEL; 26261; Arath; 3145; 26261.
DR	PROSITE; PS00375; UDPGT; 1.
DR	PFAM; PF00201; UDPGT; 1.
KW	Transferase.
SQ	SEQUENCE 456 AA; 50713 MW; 9A33B1F9 CRC32;

Result No.	Query	Score	Match	Length	DB	ID	Description
1	gb_ba1.*	1458	98.9	1458	7	AB013597	AB013597 Perilla f
2	gb_ba2.*	1329.4	90.2	1507	7	AB013596	AB013596 Perilla f
3	gb_om.*	680	46.1	2043	7	AB013598	AB013598 Verbena h
4	gb_ov.*	439.8	29.8	1864	7	AB027455	AB027455 Petunia x
5	gb_pat.*	284.4	19.3	19894.4	8	ATCHRIV38	AL161538 Arabidops
6	gb_ph.*	284.4	19.3	200576	7	ATFCA0	297335 Arabidops
7	gb_p11.*	262.4	17.8	1594	7	AB000623	AB000623 Nicotiana
8	gb_p12.*	226.8	15.4	84203	49	AC005106	AC005106 Genomic s
9	gb_p1.*	226.8	15.4	103223	8	AC007153	AC007153 Arabidops
10	gb_pr2.*	186	12.6	195068	8	ATCHRIV42	AL161541 Arabidops
11	gb_pr3.*	186	12.6	197419	8	ATCHRIV41	297339 Arabidops
12	gb_ro.*	186	12.6	205065	7	ATFCA4	U81293 Arabidops
13	gb_sts.*	181.2	12.3	1589	8	ATU81293	A62523 Sequence 28
14	gb_sy.*	142.4	9.7	1513	5	A62523	A72396 Sequence 28
15	gb_in1.*	142.4	9.7	1513	5	A62523	AC006533 Arabidops
16	gb_in2.*	138.8	9.4	99188	49	AC006533	A62526 Sequence 31
17	gb_ba1.*	131.8	8.9	1588	5	A62526	A62525 Sequence 30
18	gb_ba2.*	129.6	8.8	1459	5	A62525	A62529 Sequence 34
19	gb_hum3.*	126.4	8.6	1369	5	A62529	AC002333 Arabidops
20	gb_hum4.*	125.4	8.5	94786	49	AC002333	AF190634 Nicotiana
21	gb_pr4.*	124.6	8.5	1440	49	AF190634	L34847 Zea mays IA
22	gb_htg3.*	121.6	8.2	1416	7	MZEIAGLU	AB019232 Arabidops
23	gb_htg4.*	118	8.0	80818	7	AB019232	A62509 Sequence 14
24	gb_htg5.*	116.8	7.9	566	5	A62509	

Qy	354	actccacac	ctcttgc	atg99gc	g99gag	gtgc	gctttgt	ccccaa	gctccg	accgcc	ccc	413
Db	360	actccacac	ctcttgc	atg99gc	g99gag	gtgc	gctttgt	ccccaa	gctccg	accgcc	ccc	419
Qy	414	ttctctggg	tcgagcc	gcac	gctg	ctg	gtgata	ccactt	cttact	ctcaac	gcgtac	473
Db	420	tactctggat	tgagacc	agctac	gggtgg	atgata	tatttt	tacttt	tttttt	ttaac	ggcctata	479
Qy	474	cagacagat	gcagcc	gggttcc	aatga	attca	---g	ctccct	cggttcc	atccct	tcg	530
Db	480	gcacacaaa	tgcatt	gcgggtt	ccgatg	ctatt	tacttt	ccg	ggagga	ctcc	acgtgtcg	539
Qy	531	agcagcgag	ctctcc	ga	cggttct	gctgc	ctgcac	cccg	agagatt	cggttgc	atga	590
Db	540	ccacagctg	atnatt	accgctt	tcctt	ccctt	ccac	gcatt	gagagatt	ccggtt	actctga	599
Qy	591	tgaagagag	actgga	aaactt	bag	cggtga	agagag	gcg	aaagtatt	ggtg	aacacgt	650
Db	600	tgaagagag	aaatt	tggaaat	ttt	aga	aggtga	aaaaa	acc	taag	gcttgg	659
Qy	651	ttgatcggt	tgagcc	cgatc	actc	acgc	gtattga	tag	tatg	atgatt	aatcg	710
Db	660	ttgatcggt	tggagcc	ctg	atg	cgctt	ca	ggcc	ttgat	agat	gatt	719
Qy	711	ggcgttgat	ctccct	cgccctt	ctt	ggag	cgga	agatc	ccct	ccga	aa	770
Db	720	ggcgttgat	ctccct	cgccat	ttt	ggag	cggt	taaa	agatt	cc	ttt	779
Qy	771	gcgatctt	tcgaaaa	---tc	gag	agaga	ata	act	cg	tgag	gttga	827
Db	780	gagattt	gttcg	aaag	gggtc	aa	tgac	gcag	attg	ctt	cgat	839
Qy	828	cgaaatct	tcg	ggtgt	atg	tg	ctg	ttt	ggag	cg	ttt	887
Db	840	ctcgatct	tcg	ggtgt	atg	tg	ctg	ttt	ggag	cg	ttt	899
Qy	888	tgaagagat	tgga	aggctat	ag	ctc	gcga	agcc	ctttt	atg	atag	947
Db	900	tgaagagat	tagca	agggc	gtt	agat	gtg	gag	ggcc	ttt	gtg	958
Qy	948	aacagaga	atgac	gc	gcg	cgaga	aga	aga	aga	agat	gtt	1007
Db	959	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	998
Qy	1008	aattgaaaa	aat	tg	ggg	aaaa	tag	tgc	gtg	ctc	gcag	1067
Db	999	agttgaa	acag	tg	ggg	aaaa	ttg	atct	tg	gtgt	ctca	1058
Qy	1068	ctgcgtt	gggat	ttt	ctg	ag	ccatt	g	gggtg	gaact	cg	1127
Db	1059	ctcgtt	gggat	ttt	ctg	ag	ccatt	g	gggtg	gaact	cg	1118
Qy	1128	gcgggatt	ccggt	tg	gtg	cg	ggtgc	gcag	ttt	ggag	gtt	1187
Db	1119	tcggggt	tc	cgat	tg	gtg	ggctt	ttt	cc	gcag	ttg	1178
Qy	1188	ttgagat	gc	gtg	ggg	gac	gggtg	agat	tg	aat	gaa	1244
Db	1179	tggaggat	gtg	tgc	gag	ac	gggtg	tgcag	atg	ta	gag	1238
Qy	1245	atggat	gtag	at	aga	ag	gtgtg	t	ggag	at	ggat	1304
Db	1239	atggat	gtat	ga	aa	tt	agag	at	ggat	tt	ggat	1298
Qy	1305	tag	tga	gag	aaaa	tg	ccat	tc	aa	atg	ga	1364
Db	1299	aact	tag	ag	ag	at	gtc	gga	ag	att	tg	1358
Qy	1365	gat	ctt	act	ca	aa	at	ct	ca	ag	ctt	1424
Db	1359	gat	ctt	act	ca	aa	at	ct	ca	ag	ctt	1418

Qy	1425	aaacag	1430
Db	1419	AATGAG	1424

RESULT 4					
LOCUS	AB027455	1864 bp	mRNA	PLN	07-JAN-2000
DEFINITION	Petunia x hybrida PHI mRNA for anthocyanin 5-O-glucosyltransferase, complete cds.				
ACCESSION	AB027455				
VERSION	AB027455.1	GI:6683051			
KEYWORDS	anthocyanin 5-O-glucosyltransferase.				
SOURCE	Petunia x hybrida	(cultivar:Surfinia)	corolla	cdna	to mRNA.
ORGANISM	Petunia x hybrida				

REFERENCE					
AUTHORS	Yamazaki,M. and Saito,K.				
TITLE	Anthocyanin 5-O-glucosyltransferase				
JOURNAL	Published Only in DataBase (2000) In press				
REFERENCES	2 (bases 1 to 1864)				
AUTHORS	Yamazaki,M. and Saito,K.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences; Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan (E-mail:mamiy@p.chiba-u.ac.jp, Tel:81-43-290-2905, Fax:81-43-290-2905)				

FEATURES					
source					
1..1864					
/organism="Petunia x hybrida"					
/cultivar="Surfinia"					
/db_xref="taxon:4102"					
/tissue_type="corolla"					
100..1506					
/gene="PHI"					
100..1506					
/gene="PHI"					
/codon_start=1					
/product="anthocyanin 5-O-glucosyltransferase"					
/protein_id="BAAB9009.1"					
/db_xref="GI:6683053"					
/translation="MQPPIHLYTTPAOGHINPALQFKNLKVKGIEVETSTSYAQS RMDSKLINAPKUNLPFFSDGEVDHSKDQPFYMSQLRKCSEIVKKILIFCSND GPOTICLLYSIFLPMAAVEAREVHIPSSALLNSQPATILDIIYFNPHGVKAMANESDN PNWSIQIPGLPLETRDLPFLPYGAKSLRALPPPEKLIDTDAETTPKILVNTP DELPEALNAITEGYKYFGIGLPLPSLAGNDPLDASFGDLFONSDYMELWLNKSP SVVVISFGSLMNPISOMBEIESKGLIDIGRPFLMWIKENKKEBENKKGTCLE KXIGVPCMSOLEVLHPVSGCVSHGWNLSALESLACVPVAFPQWTQMTNAKQV EDVMKSGVRVIRINDPVSESEIKRCILVMDGGEGRELKNKKKELAREAVRG GSSHNNLKAFIDDVAGGF"					
BASE COUNT 603 a 318 c 369 g 574 t					
ORIGIN					

Query Match 29.8%; Score 439.8; DB 7; Length 1864;					
Best Local Similarity 61.0%; Pred. No. 2.6e+86;					
Matches 865; Conservative 0; Mismatches 497; Indels 57; Gaps 7;					

Qy	25	aaaaatggtccgcgccgcgtgctgtagcaaacgtttccggcgcaaggccacataaatcc	84
Db	96	AAATATGGTGAGCCCTCATGTCTATTAAACAACATTTCCAGCACAAAGGCCATTAATACC	155
Qy	85	cgccttcattcgaagaagactcctaaagccgcgactgcagtgcaccttttcacgaq	144
Db	156	AGCACTTCANTTGCACAGAATCTTGTCAGATGGGCNTAGAGTGNCAATTTCTTCAAG	215
Qy	145	cgttatgatggcgccgatggccaacaagcctccgcgcgtgcggaaaccaccacggg	204
Db	216	CATTATTGCCAAAGCGGTATGGATGAATAATCATTCTTAATGC-----ACCAAAGG	269

QY 205 cctcgacttcgtggtctccacggtctacgacgagcggtcggaagcccggtcgacggtg 264
Db 270 ATTGAATTCATCCATTTCCGATGCGTTGATGAAGGTTTGTATCATTCATAAAGACCC 329
QY 265 gaagcgctacatgtccagatgaaagcccggtcgaggtcggttaagaacacctctct 324
Db 330 TGTATTTTACATGTGCACAACTTCGTAATGTGGAAGTGAACGTCAAAAATAAATCT 389
QY 325 caacaagcaga-----cgtaacttcgtctactccaccctcttggatg 372
Db 390 CACTTGCTCTGAAATGCAGAGCTATACTTCGCTTACTTCTCATTTTCTCTCTTG 449
QY 373 ggcggcgaggtggtcggtttgcccacgtcccgacccgcccctctctctgtgggtcgagccgc 432
Db 450 GCGACGAGGTAGCAGGTGAAGTTCACATPCCCTTCTGCTCTCTTGTGAGTCAACACGC 509
QY 433 caccgtgctgtgcataccactctactcaagctcactcagcagcagcagatcgagccgg 492
Db 510 AACAATATGGACATATATTACTTCAACTTTCATGATATGAAGAAGCTATGGCTAATGA 569
QY 493 ttccaatgaa-----attcagctccctcggttccatccctcggtgagcagcgag 540
Db 570 ATCCAATGATCAAAATTTGGTCCATTTCAACTTCCCGGGCTTCCACTACTGGAACCTCGAGA 629
QY 541 tcttcgagcgtttctgtgc-tgcgacgcggagagattccggttgatga----- 590
Db 630 TCTTCTTCAATTTTACTTCTTATGTTGTCGAAAGGAGTCTTCGAGTTGCACCTCCACC 689
QY 591 --tgaaggagaagctggaacttttagacggtggaagagaagcggaagtatttggtaaacac 648
Db 690 ATTCAAGAATTTGATAGACACATTAGATGCTGGAACCACTCTCAAGATCTTGTGAATAC 749
QY 649 gttgatcgctgtgagccgacacacggtattgatagttgatgattgagtcgggat 708
Db 750 ATTTGATGAATTAGAGCTCGAGCACTCAATGCAATTGAAGTTATAGTTTATGGAAAT 809
QY 709 cggcgctgtgattccctcgccctcttggacggtggaagatccctcgaaagctcttaagg 768
Db 810 TGGACCGTGTGATCTCTCTCTCTTCTGCGTGAATGACCCCTTTAGATGCTTCAATTTGG 869
QY 769 cggcgatctttcgaaataatcgagagataactcgtggagtggttgaaactcgaagcc 828
Db 870 TGGTGATCTTTTCAAAATTC-----AATGACTATATGAAATGTTAAATCAAGGCC 923
QY 829 gaaatcttcgggtgtatgtgctgttggagcgttttgaggtttcccaagggcaacaaat 888
Db 924 AAATTCATCAGTTGTTATATATCTTTTGGAGTCTAATGAATCCATCTATTAGCCAAAT 983
QY 889 gaaagagattgggaaaggctattagcctgcgaaagccctttttatgatgatacagaga 948
Db 984 GCAGGAGATATCAAAAGGTTGATAGACATAGGAAGCGCGTTTATGGGTGATAAAGA 1043
QY 949 acagaagaatgacagcgcgaagaagaagaagaagaagagattgagttgcattgggga 1008
Db 1044 AAATGAAAAGCC-----AAAGAAGAAGAGATAAAAAGCTTGGTTGTTATGAAGA 1094
QY 1009 attgaaaaaattgggaaaaatagtcgtggtgctgcgagttggaggttcttggcgaccc 1068
Db 1095 ATTGAAAAAATAGAAAAATAGTTCCATGGTGTTCACAACTTGAAGTTCTAAAACATCC 1154
QY 1069 tgcgttggagattcttcgacgcatgtgtgggtggaactcggctgtggaagaccttgattg 1128
Db 1155 ATCTTTAGGATGTTTGTCTCATTTGATGGAATTCAGCCCTTAGAGAGTTTAGCTTG 1214
QY 1129 cgggattccggtggtgcggtgctgagtcagtcagcagcgaatgcgaagctgat 1188
Db 1215 TGGAGTGCCAGTTGTGSCATTTCTCAATGGACAGATCAATGACAAATGCCAAACAGT 1274
QY 1189 tgagatgcgttgggggacaggggtgagatggaatgaatgaagggggtggggttgatgg 1248
Db 1275 TGAAGATGTGTGGAAAAGTGGAGTGAAGTGAATTAATGAAGATGATGTTGTTGTAAG 1334
QY 1249 atgtgagatagaagggtgtgtggaagatggtgagtggtgggtgacaaagcacaactagt 1308

Db 1335 TCAGGAATCAAAAGGTGTATTGTAATGGTAAATGGATGGAGGAGAAAGGGGAAGAATT 1394
QY 1309 gagagaaatcccatcgaatggaagcatttggccgacgaagccatggataggatggatc 1368
Db 1395 GAGAAAGAATCTAAGAATAGGAAGAAATTTGGCTAGAGAAGCTGTCAAGGAAGGTGATC 1454
QY 1369 ttcatcacaatctcaacgcttcttctctcaagttgc 1407
Db 1455 TTCACACAAGAATTTAAAGGCTTTATTGATGATGTTGC 1493
RESULT 5
ATCHRIV38/c
LOCUS ATCHRIV38 198944 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 38.
ACCESSION AL161538
VERSION AL161538.2 GI:7268110
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi
1 (bases 1 to 198944)
EU Arabidopsis sequencing, project.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:
lencke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV37 at the 5' end and an
overlap with ATCHRIV39 at the 3' end.
FEATURES
source
1. 198944
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
4845..5598
/gene="AT4g14060"
join(4845..5031,5330..5598)
/note="strong similarity to major latex protein type 1,
Arabidopsis thaliana; contains EST gb:A1994471.1, T43693,
A1099764, H36765, T43905, A1998305.1, R29829, T22625,
R90142, T41926, T22170, W43041, R65053, AN712594,
AN712597"
/codon_start=1
/product="major latex protein like"
/protein_id="CA78448.1"
/db_xref="GI:7268111"
/translation="MATSGTVTVTEVPLKSGAEKHYKSWRNENHLFPDAIGHHIGQVTV
HDSNDSHGALKIWNYYLIDCKPEMEKERRIEDDENMAVTFVGLGHEVMEQLKYDIL
QFIQKSPDDIVCKITWKEKRAADSPSPNTMKLVKLSLAANDAHVLA"
4845..5031
/gene="AT4g14060"
/number=1
5032..5329
/gene="AT4g14060"
/number=1
5330..5598
/gene="AT4g14060"
/number=2
complement(9562..9654)
/number=1
9562..15425
gene

```

CDS
/gene="AT4g14070"
complement(9562..9654,9750..9807,9883..9998,
10097..10176,10276..10435,10594..10731,11065..11145,
11247..11300,11383..11541,11653..11758,11882..11949,
12156..12305,12439..12554,12626..12834,13288..13360,
13958..14072,14674..14859,15291..15425)
/note="strong similarity to AMP-binding protein - Brassica
napus; Contains Putative AMP-binding domain signature
AA355-266; contains EST gb:AI999282.1"
/codon_start=1
/product="AMP-binding protein"
/protein_id="CAB78449.1"
/db_xref="GI:7268112"
/translacion="MASTSLGSPSTLLSYGSPSRQFPDGFRLISGHESVRIPSRFRF
REKVPKSSPFLSSFSFGDAALRSSEKAVPDWRSSAEKYGDRVALVDPYHDPPLK
LTKQLQELTDLAEGLRLVGLVKADEKIALPADNSRWLVSDQGVNVRGSRVVE
LLOIYRSESAVVDNPEFNRIASFSTKASLRFLLLWGEKSSLVQGMQIPVYS
HQIILNGQSRKLSASNDTSIRNOFIDSDTAAMTSGTTGNPKGVMLTHRNLL
AEIKHLSKYPAQAGDKFLMPLSNHAYERASEYFIFTCGVEQMYTSIRYKDDLRKY
QNTYIVSVLYVETLYSGTKQTSASSAGKFLALTLIKVSMAYEMKRIYEGCLTK
HVDKFFAIGVILONGVGLTETSPVVCARTLSCNVLGSAGHPMGTEFKIVDPPTNV
LPPGSKIIKVRGPVMKGYKYNPSTTKOVLNCSGWENTGDTGMIAPHKSGSRHCG
GVIVLEGKADTIVLSTNGENVEPLELEEAAMBSRVIEQIVIGODRRELGAIIIPKE
BAQVDPETSKETLKSILVQELRKKTSECSFOVGVLVDPPFTIDNGLMTPTKIRR
DMVYAKKEIDQLIS"
complement(9655..9749)
/number=1
complement(9750..9807)
/number=2
complement(9808..9882)
/number=2
complement(9883..9998)
/number=3
complement(9999..10096)
/number=3
complement(10097..10176)
/number=4
complement(10177..10275)
/number=4
complement(10276..10435)
/number=5
complement(10436..10593)
/number=5
complement(10594..10731)
/number=6
complement(10732..11064)
/number=6
complement(11065..11145)
/number=7
complement(11146..11246)
/number=7
complement(11247..11300)
/number=8
complement(11301..11382)
/number=8
complement(11383..11541)
/number=9
complement(11542..11652)
/number=9
complement(11653..11758)
/number=10
complement(11759..11881)
/number=10
complement(11882..11949)
/number=11
complement(11950..12155)
/number=11
complement(12156..12305)
/number=12
complement(12306..12438)
/number=12
complement(12439..12554)
/number=13
complement(12555..12625)
/number=13
complement(12626..12834)
/number=14
complement(12835..13287)
/number=14
complement(13288..13360)
/number=15
complement(13361..13957)
/number=15
complement(13958..14072)
/number=16
complement(14073..14673)
/number=16
complement(14674..14859)
/number=17
complement(14860..15290)
/number=17
complement(15291..15425)
/number=18
complement(16137..16174)
/number=1
16137..17732
/gene="AT4g14080"
complement(join(16137..16174,16252..17604,17687..17732))
/note="the cDNA A26668 is cognate to this gene; gene A6
for anther-specific protein - beta-1, 3-glucanase homolog;
Contains Glycosyl hydrolases family 17 signature
AA274-287; Prokaryotic membrane lipoprotein lipid
attachment site AA8-18; contains EST gb:AW004339.1,
218400"
/codon_start=1
/product="A6 anther-specific protein"
/protein_id="CAB78450.1"
/db_xref="GI:7268113"
/translacion="MSLLAFLEFTILVFSSSCSAFRQGHRYMORKTMDLASKIGI
NYGRGNLSPYQISINFIKIKAGHVKLYDADPSLTLLSQTLYVTVPNHQITA
LSSNOTIADWVRTNIPYQTOIRFVLVGNELISYNSGVSNVLPAMKRYNLSLR
LHG1HNIKVGTPLAMDSLRSSFPSTPREETGPVMLPLKELNGTNSYFELNVHP
YFRSRPMNTSLDFALFQGHSTYTDQGLYRNLDDMLDSVLFAATKLGYPHML
AISEIWNFGDIDETGANTNATNRNLKAKMSAPFIPSPRGDLPITFVFSLF
NENKSGSTGRHWGLIHDPGSDIYDVFQTPPLGFNPLPKPNTNNVYKGVWCVP
VEGANLELEETLRMACAQSNTTCAALAPGRCYEPVSIYHASYALNSYNAQFRNQ
IQCFNGIAHETTTNPGNDRCKFPSTL"
complement(16175..16251)
/number=1
complement(16252..17604)
/number=2
complement(17605..17686)
/number=2
complement(17687..17732)
/number=3
18189..18494
/gene="AT4g14090"
/number=1
18189..25761
/gene="AT4g14090"
join(18189..18494,18694..18787,18938..19018,19571..19724,
19772..19833,21489..21524,22014..22792,22844..22948,
23060..23200,23280..23566,24375..24631,24686..25140,
25192..25311,25445..25614,25665..25761)
/gene="AT4g14090"
/note="similarity to proton-coupled peptide transporter
PEPT, Rattus norvegicus; Contains Prokaryotic membrane
lipoprotein lipid attachment site AA535-545; Prokaryotic
membrane lipoprotein lipid attachment site AA955-965;
contains EST gb:AI998193.1"

Query Match 19.3%; Score 284.4; DB 8; Length 198944;
Best Local Similarity 56.2%; Pred. No. 3:3e-52;
Matches 773; Conservative 0; Mismatches 51; Indels 92; Gaps 9;

```


[illegible]

misc_feature	/number=5	20514..20606	/note="EM_ORG:CHSOPSBO Spinach plastid pabb operon with
Query Match	19.3%	Score 284.4; DB 7; Length 200576;	
Best Local Similarity	56.2%	Pred. No. 3.3e-52;	
Matches	773;	Conservative 0; Mismatches 511; Indels 92; Gaps 9;	
QY 48	tgtatgaacggtttccgagcgaagccacataaaatcccgccctccaatccgcgaagagac	107	
DB 81653	TGCTTTAAATATCCGACGGAAGGTCAATCAACCGCGGCTTCAACTAGCCAAACCGCC	81594	
QY 108	tctataaagccggcactgacgtccagctttttccacagcgttttatgcatgcgccgcatg	167	
DB 81593	TCATCCACACGGTGCACACCTGCATATCCACCGCAGTCTCTGCTCACCGACGTATGG	81534	
QY 168	caaacacagcctccgcccgtgcggaacccacccgcccctcgcactctggtggtctccg	227	
DB 81533	GCAGACCACTTCCA-----CAAAAGGTCTATCTTTCGCTTGGTTTCACCG	81489	
QY 228	acggtcaacacagcgggtgaagcccgccgagcggagcgcgtcatgctccgagatga	287	
DB 81488	ATGATTCACGACGGTCTCAAGTCATTGGAAGACCAAGAAATCTACATGTCGGAATCA	81429	
QY 288	aagcccgccggtccgagccttaagaacccctctctcaaacagcagcagc	338	
DB 81428	AACGATGTGTTTCAAAACGCCCTGAGAGACATCATCAAAAGCCAAATCTTGACGCCACACCG	81369	
QY 339	-----tcacatttgctgctactccactccactcttctgcatggcgccggggtgg	386	
DB 81368	AAACAGAGCTATCACCGGGTAAATCTACTGTCTCTCGTCCGCTGGTTCCTACGGTAG	81309	
QY 387	cgggttgctccacgtccgacccctctctctggtcgagccgcgcacgctgctgtga	446	
DB 81308	CGGTGAGTTTCACTCCCACTACACTCTCTGATGTAACCACTGCTGCTACTAGACA	81249	
QY 447	tataccactctactcaacggctacgcagcagagatgcagccggttccaatgaattc	506	
DB 81248	TCFACACTACTACTTCAACACCTCTTAAACACTCTCTCGAGCT---GAACCCGATTA	81192	
QY 507	agctccctcggtctccatccctcgagcagcagctctccgagcttctgctgcctga	566	
DB 81191	AATACCGAATGTCACCTGATCACCCGGTGACCTCCGCTGCTTCTTCAACCTTCA	81132	
QY 567	cggc---gagagattccggttgatgataagggagagctggaaactttagacggtga	623	
DB 81131	AGCATATACCGTCACTGCTGTGATCTTAAGAGAACATATCGAAGCTCTCGAAGCGAAT	81072	
QY 624	agaagcgcaagtattggtgaacacgtttgagcgttggtggagcccgatgactcacgcta	683	
DB 81071	CARACCTTAAGATTCTGTAAACATCTCTGCTTTGGAACAGATGCTTTAACTCTG	81012	
QY 684	tgatgggtatgattgattcggtgattcgggccggttgattccctccgcttcttggacgg	743	
DB 81011	TTGAGAAATCAAGATGATGCCAATCGACCGTGGTGTCTTCTCCGAGGGTAA-----	80955	
QY 744	aagatccctccgaaacgtcttacgagcggcgatcttttcgaaaaatcgaggagaaat	803	
DB 80956	-----AACCGATCTTTTCAATCTTCCGACGAGAT--T	80925	
QY 804	gcgtgagtggttgaaactcgaagccgaaatctcggtgggtgattgctgcttggga---	860	
DB 80924	ACAGAAATGGTATGATCTGAAGCTCGAGAGATCAGTGATTTACATTTCTTATAGGCAC	80865	
QY 861	gcgttttgaggtttccaaaggcacaaaaggagagattgggaagggcattagcctgcg	920	
DB 80864	ACCCGATGATTTACAGAGAAACATGGAAGCGCTTACTCACGGCGTGTAGCTACAA	80805	
QY 921	gaagggcccttttatgattgatacagagaacagaagaatgacagcgcggaagaagaag	980	
DB 80804	ACAGACCGTTTTTATGATGCTGAGGAGAGAAAATCCAGAG-----AGAA	80758	
QY 981	aagaagaagattgattgattggggaattgaaaaaaatgggga--aaatagtgctgctg	1039	

ALLENALRVLNKNTVWIGLIPSPFLDPKDRKDNFFAADMIESNNYMEWLDARANK
SVYIARCSVAETSSQRMREISGGLLKCRPFELWVIRETLNGRPFEBKLTCKDELEKI
GRIYWCQSMELVKHSLGCFUFGNGSNFLESASGVPVACPIWMDQICNAKLIQD
VMKIGVRYNANKRGIIIKRBFQKCIIEVNGDAEAGEELKNAQKWKDLAKESTKENS
SNVMLKAYVNECLLHGHC*
1594

polya_site

BASE COUNT 514 a 238 c 325 g 517 t
ORIGIN

Query Match 17.8%; Score 262.4; DB 7; Length 1594;
Best Local Similarity 59.0%; Pred. No. 1.9e-47;
Matches 496; Conservative 0; Mismatches 326; Indels 18; Gaps 2;

QY	590	atgaaggaagcctggaacatttagacggtgaaggaagcgaagatttggtagaacgcg	649
Db	540	ATTAAAGCAATTTGAGATATTGANTAGTGAAGAAATCCAGAAATTTCTGTGAATACT	599
QY	650	tttgatgcttggagccgccttctggaagcgaagatccctccgaacgtcttacgac	709
Db	700	TTTGATGCAATTTGAACTTAATGCTTTAGGCGTTTGAATAATTAACATATGCTTGAATT	759
QY	710	gggcggtgattccctccgcttctggaagcgaagatccctccgaacgtcttacgac	769
Db	760	GGACCTTAATTTCTTCATCTTTCTTTGATGAGAAAGATCAATTTTTCGGG	819
QY	770	ggcgatcttttcgaaaaatcgagaggaataactgcggtgagtggttgaaactcgaagccg	829
Db	820	GCTGATATGATCGAAA-----GTGAAATAATTTACATGGAGTGTGGATGCAAGGGCG	873
QY	830	aaatctcggttggtatgtctgtcttggagcgttttgaggtttccaaagccacaaatg	889
Db	874	ATAAATACGATTATTACATAGCAATTTGGTAGCTATGCTGAGATATCAAGTCAATGGATG	933
QY	890	gaagatattgggaaggcctattagcctcgggaagcgcccttttttggatgacagaaa	949
Db	934	GAAGAGATTAGTCAAGGATTGTTGAAATGTTGGAAGACCAATTTTGTGGTGATTAGAGAA	993
QY	950	cagaagaatgacgcgcgcgaagaagaagaagaagaagagtggtgagcttgggga	1009
Db	994	ACCCFANAT-----GGGNAANACCCGAGGAAAGTTGACTTTGAAGATGAA	1041
QY	1010	ttgaaaaaatggggaataatagtcgtcgtcgcagcttgaggtttcttgcgcacccct	1069
Db	1042	TTGGAAGAGATAGGAGAGATAGTCGCTGTGCTCGCAATTTGGAAGTCTCTAAACACTCG	1101
QY	1070	gcgttggaatgttcgtgacgcattgtggttgaaactcggctgtgagagccttgagttgc	1129
Db	1102	TCCGTAGGCTGTTCTTGACTATTTGGATGGAATTCGACTCTGGAGAGCTTAGCTTCT	1161
QY	1130	gggattcccggttggtgcggtgcccagtggtttgatacagcgcgaatcggaagctgatt	1189
Db	1162	GTGTGCCAATCGTGGCATGCCGATGCGCAATTTGGAATGATCAAAATTTGCAACGCAAGCTCAT	1221
QY	1190	gggattcggtggggacaggggtgagatgagaatgaataagggggtgggttgatgga	1249
Db	1222	CAAGATGTTTGAAGATTTGGTGTGAGAGTGAACCTTAATAAAGAGTATCATTAAGA	1281
QY	1250	tgtgagatagaaggtgtgtgagatggtgatggatgggggtgacaaagcacaactagtg	1309
Db	1282	GACGAGTTCCAGAGTGTATAGATCGTAATGGGAGATGCTGAGGAAGAGAGGAATTG	1341
QY	1310	agaaaaatgcataaagaaagcatttgccagacagcagcagcagcagcagcagcagcagc	1369
Db	1342	AGAAAGATGCCAAAATGAAGAGATTTGGCTAAGGAATCTACGAAGGAATAATAGTCA	1401
QY	1370	tactcaacaactcaaacgcctttcttcgtcaagtgtgcaacttttaactctgctcaaca	1429
Db	1402	TCTAATGTAATCTTAAGGCTTATGTTAATGAATGTTTACTTGTGCTATTGTTAAGAAAA	1461

RESULT 8

AC005106/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT
FEATURES
source

CDS

AC005106 84203 bp DNA PLN 16-DEC-1999
Genomic sequence for Arabidopsis thaliana BAC T25N20 from
Chromosome 1, complete sequence.
AC005106
AC005106.2 GI:6587719
HTG
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosida II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 84203) Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H.,
Shinn, P., Buehler, E., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M.,
Conway, A., Conway, A., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M.,
Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and
Ecker, J. R.
Genomic sequence for Arabidopsis thaliana BAC T25N20 from
Chromosome 1
Unpublished
2 (bases 1 to 84203)
Ecker, J. R.
Direct Submission
Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 84203)
Ecker, J. R.
Direct Submission
Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 84203)
Ecker, J. R.
Direct Submission
Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
5 (bases 1 to 84203)
Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C.,
Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chlou, J., Choi, E.,
Cond, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, I.,
Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S.,
Mukharasy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J.,
Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R.,
Federspiel, N., Theologis, A. and Ecker, J.
Direct Submission
Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Dec 16, 1999 this sequence version replaced gi:3366536.
Location/Qualifiers
1. 84203
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(join(1491..2159,2389..2452,2568..2703,
3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))
/note="5' end of hypothetical protein YUP8H12.1; overlaps
T7 end of YUP8H12"
/codon_start=1
/evidence=not_experimental
/product="T25N20.1"
/protein_id="AAC80580.1"
/db_xref="GI:3935137"

Db 69760 TTTCTTACTTTCCTCCGACGGATTTCGACGACGGAGTCACTTCCAACACCGACGAGTCCA 69701
QY 262 cgggaagcgtacatgtccgagatgaagccgcggctccgagggccttaagaacacctoct 321
Db 69700 AAACCGGTGTGTACTCTCGAAGCTAATGCGGATAAAGCTCTATCGGATTTTCATCGAAGC 69641
QY 322 lctcaacaacgaaca-----cgctacattctgctctactceacacctctttgcatgggc 375
Db 69640 TAATCAGAAATGGTCACTCTCCGTAAGTTCCTTGATGTACAGATTCATCCCAACTCGGT 69581
QY 376 ggcggaggtggcggctgtgtccacgctcccgacgcccctctctctgggtcgagccgcac 435
Db 69580 TCCAAAGTGGCGGTAGATTATCTTCCTCTGTTCACTCTCTGATCCTCAACACCGCTT 69521
QY 436 cgtgctgtgcatacaccacttctacttcaagcgtctacgacgagatcgacgcccgttc 495
Db 69520 CGCTTTCACATTTATTACAATTAATCTACAGGAACAACACTCGCTTTTCG----- 69469
QY 496 caatgaattcagctccctcgcttccatccctggagcgcagctcttcogagcttct 555
Db 69470 -----AGTTCCGGAATCTACCTCTCTCGAAATCCGCGATCTGCTTCTTCT 69422
QY 556 gctgctgcgacgcgcgagatctc---cggttgatgatgaaggaagcctgaaacttt 612
Db 69421 CTCACCTTCCACAGACGACAAAGCCGACAAAGCAGTATATCAAGAACTGATGATTTCT 69362
QY 613 agacggtgaagagagcgaatattgttgaaacacgcttggatcggtggagccgcatgc 672
Db 69361 CAAAGAAGAAATCTAACCCGAAATTCCTGTCACACATTCGATTCGCTGGAGCCAGATT 69302
QY 673 actcagcgtattgataggtatgattgatcggtatcggttcggttgatccctcgccct 732
Db 69301 CTTAACACTATTCTGAATATAGAAATGGTGGCAGTGTGCTCTTACTTTCCTGGGAGAT 69242
QY 733 cttggacgcgaagatccctccgaacgcttaccgcgcgcatctttccgaataatcga 792
Db 69241 TTT-----CACTGGAAGCAATCAGGTAAAGATTTTCAAGAGATCATCA 69197
QY 793 ggagataactgcgtggagtggttgactcgaagcgaatcttcggtgggtatgtctc 852
Db 69196 AAGTAGTATTATACACTTTGGTTAGTCTCGAAGACAGAGTCTCTGTTATTATGTTTC 69137
QY 853 gtttgggagcgttttgaggtttccaaaggccacaaatgggaagatgggaaaggcctatt 912
Db 69136 TTTTGGAAACAATGGTTGAGTTTTCGAAGAACAACAGATAGAGAACTAGCGAGACACTCAT 69077
QY 913 agcctgcggaagcccttttatgattgatcagacgaagaagaatgacgacgcgcaaga 972
Db 69076 AGAAGGGGCAAGCGTCTCTTGGGTATTAACTGATTAACTCACAGACGAGCGAATAAT 69017
QY 973 agaagaagaagaagaagattga-----gttgcattgggggaattgaaaaa 1017
Db 69016 AGAAGGAGAAGAAGACAGACAGATTGAGAAGATAGCTGTTTATAGACAGAGCTTGAAGA 68957
QY 1018 aatggggaataatgtcgtggtcgcagttggagttctggcgacccctgcgttggg 1077
Db 68956 GTTTGGGATGATTTGCTCGTGGTGTCCAGATAGAGGTTTTCACACACCGACCATAGG 68897
QY 1078 atgttctgacgcattggtgggtggaactcggctgtggagagcttgagttgcggattccc 1137
Db 68896 TTGTTTTTTGACTATTGTGGGTGGAGCTCATCATCTGGAGAGTTTGGTTCTCGCGTTC 68837
QY 1138 gttgctgctgcgtccgcagtggtttgatcagacgacgaatcgaaagctgattgagatgc 1197
Db 68836 AGTGTGGCGTTTCGATGTGGTGGGATACCCAGCAATATCGAAGCTTTTTCGAAGAAAT 68777
QY 1198 gtgggggacaggggtgagagtgagaatgaatgaaggggggtgggggttgatgtgagat 1257
Db 68776 ATGGAAGACAGGTGTGAGGGTGTAGAGAACTCGGAAGGTTTGTAGTAGAGAGGAGAT 68717
QY 1258 aaaaaggtgtgagatggtgattgattgggggtgacaagaccacaaactagtgagaaaaa 1317

Db 68716 AATGCGGTGTTTGGAACACGATGATGAGGCG-----AATTCGGTGGAGCTGAGGAAAA 68663
QY 1318 tgccatcaaatggaagacttttggccagacaagccaatgggatagatggatctctcactcaa 1377
Db 68662 CGCAGAGAATGGAAGCGCTTTAGCGACTGAAGCGGGTAGAGAGGAGGATCTTCGACAA 68603
QY 1378 caatcacaacqccctttct 1395
Db 68602 GAATGCGAAGCTTTGT 68585
RESULT 9
AC007153/LOCUS AC007153 103223 bp DNA PLN 17-MAY-1999
DEFINITION Arabidopsis thaliana chromosome I BAC F3F20 genomic sequence,
complete sequence.
AC007153
VERSION AC007153.2 GI:4580365
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi.
1 (bases 1 to 103223)
Fedorpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Unpublished
2 (bases 1 to 103223)
Fedorpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (24-MAR-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 103223)
Fedorpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (08-APR-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 103223)
Fedorpiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT
On Apr 9, 1999 this sequence version replaced gi:458084.
Bases 1-33654 of clone F3F20 overlap with bases
50550-84203 of 'TAMU' clone T25N20, gb|AC005106.
e-mail for correspondence: arabsequence.stanford.edu
Genes with similarity to proteins in the databases are described as
'putative', '-like' or 'similar to'. Genes that have EST
similarity, but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction software are described as 'hypothetical proteins'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory).

gene	/db_xref="GI:7268304"		intron	/translation="MASSAFAPPSYIITKGLSTDSCKSTSLSSRSRLVTDLPSPCLK PNNSHNRRAKVCASLAKEGEYYSNRPTPLDITINYPIMKNLSYKELKQLSDELK SDYIFNVTGHLGSSLGVELTVALHYIFNTPODKILMDVGHQSYPHRIILTGRGK MPTMRTNGLSGFTKRGSEHDCFTGHSSTISAGLCMAVGRDLKGNNNVAVIGD GAMTAQAYEAMNAGYLDSDMIVLNDKNQSLPTATLDGSPPPVGCALSSALSRLQS NPALRELRVAKMGTQIGGPHQLAAKVDEYARGMSLGTGSSILFEELGLYIYGPVDG HNTDDLVLKGVKSPRTGPVLJHVTEKRGYVYAEERADDKYHGVVKFDPATGRQF KTNTKQSTTTFAEALVAEYDQVVAIHANMGSGTGLNLFQRPPTRCFDVGIAE QVATYFAAGLEGKLPCCALITYSFMQRAYQVVDVLDLQPLVPFAMDRAGLVGADG PTHCAPDVTYFACLPNMIVMAPSDEADLFNNVATAVAIDRPSCFRYPGNGIGVAL PPNGKVPYIEIGKRGRLKEGERVALLGYSVQSLGAAVMEERGLNVTVAADARCK PLDRALIRLSAKSHVELITVERGSGIGFGSHVQVFLALDGLLDGKLRWPNVLPDRYI DHGAPADLAEAAGLPSHIAATALNLIGAPREALF"	
	9917. .12953			/gene="AT4g15560"	
intron	10067. .10349		exon	/gene="AT4g15560"	
	/number=1			10350. .10475	
intron	10476. .10558		exon	/gene="AT4g15560"	
	/number=2			/gene="AT4g15560"	
intron	10559. .10874		exon	/number=2	
	/gene="AT4g15560"			10875. .10980	
intron	/number=3		exon	/gene="AT4g15560"	
	10981. .11252			/number=3	
intron	/gene="AT4g15560"		exon	11253. .11339	
	/number=4			/gene="AT4g15560"	
intron	11340. .11613		exon	/number=4	
	/gene="AT4g15560"			11614. .11690	
intron	/number=5		exon	/gene="AT4g15560"	
	11691. .11998			/number=5	
intron	/gene="AT4g15560"		exon	11999. .12080	
	/number=6			/gene="AT4g15560"	
intron	12081. .12362		exon	/number=6	
	/gene="AT4g15560"			12363. .12440	
intron	/number=7		exon	/gene="AT4g15560"	
	12441. .12725			/number=7	
intron	/gene="AT4g15560"		exon	12726. .12812	
	/number=8			/gene="AT4g15560"	
intron	12813. .12953		exon	/number=8	
	/gene="AT4g15560"			/number=9	
repeat_region	14413. .14436		gene	/number=9	
	20338. .24557			/gene="AT4g15570"	
exon	/gene="AT4g15570"		exon	20338. .20373	
	/number=1			/gene="AT4g15570"	
CDS	join(<20338. .20373,20767. .20852,20962. .21012,21256. .21367, 21467. .21572,21653. .21750,21989. .22117,22182. .22274, 22625. .22678,22957. .23015,23122. .23206,23286. .23765, 23765. .23765,23765. .237				

24183. .24269,24366. .24557)
/gene="AT4g15570"
/note="similarity to SEN1 protein, yeast; Contains
ATP/GTP-binding site motif A (P-loop) AAL3-20; contains
EST gb:NG5040"
/codon_start=1
/product="SEN1 like protein"
/protein_id="CA878599.1"
/db_xref="GI:7268305"
/translation="VGLSRKSFVLQGGPPGKQTILSIILGALMHATPARVQSKGTD
HEYKRIQMTIOEKYNHMRASPWILGVNPRDAIMPEDGDDGFEPTSGNELKPEVWNA
SRYRULRVKPSNSALDEIVLRLLSGLRDENAQTTPKIVRIGLKAHHSVASL
DLHVAQRGSAIDIKPQGTGTDIDISIRTALEAAIWMGFNFRINTVATLSFGSAL
LAKNSGFQDVIIDEAAQAVEPATLIPLATRKQVFLVDPKQLPATVISTVAQDSGY
GTSMEERLOKAGYPVKMKTQYRMHPELIPISOTPKPIVLSLITPLMASYFTLQITF
LTENQIRSEPSKQFVEGALEDGSDIEAQTTRDMHKYRGCPGCFDIHEGKESQHGA
TGSRYNLDEVEFVLLIYHRLVITWYFELKSSQLALISYINQVKTFRDKRKEFGTEA
EKVVDINTVDFGVGSAATLKSDPLMKNLIESAEQRNRLEKVKRPLNPFSEENLET
MKLUTEDMEIPDAPLYEDESLEPVAPYGGDDDFGDDGADDDVAMAGED"

[illegible]

gene 10065..10511
CDS join(12306..12597,12333..13495,13554..13835,13944..14000,
15045..15290,15396..15802,16095..16280,16380..16453,
/gene="dl3775w"
/note="similar to hypothetical protein C02F5.7,
Caenorhabditis elegans; contains EST
gb11217311:93779;315469"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10325.1"
/db_xref="GI:2244904"
/db_xref="SP:PREMBL:023399"
/translation="MDKPYGASVMSGIDRSAGRIDLEMGVGEATLYPGLSVGENO
LRGFRKRYVGIILSAOLLTLISAVVNLNPPVNDLLTSGPIILFLICIVPFIILSPF
FVSEYLDVFLHSRTSAIYFMSYLLTRFAHWIDYWFNFVWIPLIHYHOKHPVNL
LLALFVSVSGVSCAMIEGCIPLNPSYCHKEQNLPEIWNKACKSCVFNNLL
FLLFRGIVRIVQAAILLILSVVSLTAYTFAWKKKDFSLGPIILFTSLIILVTS
FIQHFPLGTSVAVYGGFSALKRKRGRDSSPSSKKKLTDRKTHSGAENVESSLT
DTGLTANGLFPPRIENSLIWCNPNVSSVGLASLAQKCTSLKSLDLQCYVGDQGLAAV
GKFCQLEELNRFCEGLTDVGVIDLWVCCSKSLKSGVAASAKITDLSLEAVGSHCK
LLEVLYLDSEYIHDKGLIAVAGCHRLNKLQCVSVTDVAFVAVGELCTSLERLALY
SFQETDKGNRAIGKSGKLLKDLTLSDCYFVSCGLEAIHNGKELRVEINGHNIG
TRGEATGKSCPLKELALLYCORIGNSALQELGKGLKAGTFDHKFQNIQDMPLEL
GEGCPMLKDLVLSCHCHITDNLNHLVQKLLFETCHMVYCPGITSAGVATVVSCEPH
IKAVLIEKNAVETERTRRRAGSVISYLCMDL"
12306..12597
/gene="dl3775w"
/number=1
gene 12306..17346
intron 12598..13232
/gene="dl3775w"
/number=1
exon 13233..13495
/gene="dl3775w"
/number=2
intron 13496..13553
/gene="dl3775w"
/number=2
exon 13554..13835
/gene="dl3775w"
/number=3
intron 13836..13943
/gene="dl3775w"
/number=3
exon 13944..14000
/gene="dl3775w"
/number=4
intron 14001..15044
/gene="dl3775w"
/number=4
source 14700..47605
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="cosmid YUP5B11.IB112/YUP5B11.IB1A8"
exon 15045..15290
/gene="dl3775w"
/number=5
intron 15291..15395
/gene="dl3775w"
/number=5
exon 15396..15802
/gene="dl3775w"
/number=6
intron 15803..16094
/gene="dl3775w"
/number=6
exon 16095..16280
/gene="dl3775w"
/number=7

Query Match 12.6%; Score 186; DB 7; Length 205065;
Best Local Similarity 59.2%; Pred. No. 1.2e-30;
Matches 362; Conservative 0; Mismatches 235; Indels 15; Gaps 2;
QY 799 taactcgctgggtggtgaactcgaagccgaatacttcgggtggtgatatgctcgttgg 858
Db 47149 TCAATACATAGAGTGTGGATATAAGCGGATTCGTCGTGCTTTATGTTTCGTCGG 47090
QY 859 gaggcttttgagggtttccaaaggcacaatacgaagagattcgggaaagggtcttaccgctg 918
Db 47089 GACGCTTCCGCTGTGGACAAAGAACACAGCTGTGGAGCTTTGTAAAGCGCTTATACAAAG 47030
QY 919 cggaaaggccctttttatggatgacagagacag-----aagaatgacagcagcgga 969
Db 47029 TCGGAGACCATCTTGTGGGTGATTACGGATAAGCTACAGAAATAAAGAGATGAGCA 46970
QY 970 aagaagaagaagaagaagagtgattgcatcattgggggaatgaaaaaatggggaatat 1029
Db 46969 AGAAGAAGGAAGAAGATTCATAGTAGTTCAGAGAGAGCTCGATGAGATAGGAATGGT 46910
QY 1030 agtgcgtggtgctgcagttggaggttctcggcgaccctcgcttgggatttctcgtgac 1089
Db 46909 GCTTTCATGGTGTGATCAGTTTACGGTTTGAATCATAGATCGATAGTGTGTCGTGAC 46850
QY 1090 gcattggtggggaactcggctgaggagagcttgagttcgggattcccggttggcgagt 1149
Db 46849 GCATTGCGGGTGGAACTCTACGCTGGAGAGCTTGGTTTCAGGAGTTCCGGTGGCGCTT 46790
QY 1150 ggcgcagtggtttgatcagacgacgaatcgaagctgattgagagatcgctgggggacagg 1209
Db 46789 TCCGAATGGAATGATGATGATGAACGCGAAGCTTTTAGAAGATTGTTGGAACACAGG 46730
QY 1210 ggtgagagt-----gagaatgaatgaaggggtgggggttgatgagatgagaag 1263
Db 46729 TGTAAAGCTGATGGAGAGAGAGAGAGAGAGTGTGGTGGTGGATGATGAGGAGAT 46670
QY 1264 gctgtggagatggtgattgggtggagacgaagaccacaaactagtgagagaaaaagccat 1323
Db 46669 ACGGCGGTGCATTGAGGAAGTTATGGAAGACAGCGGAGGAGCTTTAGAGGAATGCCAC 46610
QY 1324 caaatggaagaccttggccagacgaagccatgggagataggatgatatcttctactcaaatct 1383
Db 46609 GAGGTGGAAGGATTAGCGCGGAGGCTGTGAGAGAGAGGAGGCTCTTCTTTAATCATCT 46550
QY 1384 caacgcctttct 1395
Db 46549 CAAAGCTTTGT 46538
RESULT 13
ATU81293 1589 bp mRNA PLN 02-JUN-1997
LOCUS Arabidopsis thaliana UDP-glucose:indole-3-acetate
DEFINITION beta-D-glucosyltransferase (laglu) mRNA, complete cds.
ACCESSION U81293
VERSION U81293.1 GI:2149126
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (bases 1 to 1589)
Graham, R.A. and Thornburg, R.W.
DNA Sequence of UDP Glucose:Indole-3-acetate
Beta-D-Glucosyltransferase from Arabidopsis thaliana (Accession No.
U81293) (PGR97-044)
JOURNAL Plant Physiol. 113, 1004 (1997)
REFERENCE
2 (bases 1 to 1589)
Thornburg, R.W. and Graham, R.A.
JOURNAL Direct Submission
Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State

University, Ames, IA 50011, USA

FEATURES
source 1..1589
Location/Qualifiers
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
26..1589
/gene="laglu"
26..1450
/gene="laglu"
/EC_number="2.4.1.121"
/function="transfers the glucose moiety from UDP-glucose
to indole-3-acetic acid to form ester linked IAA-6lc"
/note="IAA-glucose synthase"
/codon_start=1
/product="UDP-glucose:indole-3-acetate
beta-D-glucosyltransferase"
/protein_id="AAB58497.1"
/db_xref="GI:2149127"
/translation="MANNSNSPTGPHELFVTFPQGHINPSLAKRLAGTISGARV
TFAASIGAYNRMFSTENPEYTLITATYSDGHDGFKSSASDKSRDQATGNMWSMR
RRGETUTELLIDNRKONRPFVCVVTILTTLVAELAREPHLPALLAWQVPTVFISF
VHYVNGEDIASMANTPSSSILKPSLPLITVROIPSPFISSNVYAFLLPAFREQIDS
LKEINKKILINTFQLEPEAMSSVPDNFKIVPGVPLLTLRDSSRGRTIEMLDTKA
DSSVVSFTGTAVLSKKQLVCEKALIQSRPFLWITDKSKYKNKEDEQEKEDICIS
SSEKSFYIGMYSCWDQFRLVNLHRSIGCFVTHCGNWSLTSVSYGVPVYAFQWQND
MNAKLLEDCKWCTGVRMKEKKEEGVVVVDEETRRCTIEYMEDKAEFEFRGNATRWK
LAAAVAREGSGSSENLKAFVDEHM"
1568..1575
polyA_signal /gene="laglu"
1588..1589
polyA_site /gene="laglu"
/note="24 A nucleotides"
BASE COUNT 433 a 342 c 375 g 439 t
ORIGIN

Query Match	12.38;	Score 181.2;	DB 8;	Length 1589;
Best Local Similarity	58.7%;	Pred. No. 1.2e-29;		
Matches 359;	Conservative 0;	Mismatches 238;	Indels 15;	Gaps 2;
QY	799	taactgcgtggagtggttgaactcgaaagccgaaattcttcggtgggtgatgtgtcgtttgg	858	
Db	823	TGNAATACATAGAGTGGTTGGATATAAAGCGGATTCGTCTGTGCTTTATCTTTCGTCGG	882	
QY	859	gagcgcttttgaggtttccaaagggccacaaaatggaaagagattgggaaagggctattagccgtg	918	
Db	883	GACGCTTGCCTGGTTTCAGCAAGAACAACAGCTTGTGGAGCTTTGTAAGAGCGTTGATCAAAAG	942	
QY	919	cggaaagccccttttcatggatgatacagagaacag-----aagaatgacgacgagcgga	969	
Db	943	TCGGAGACCACTTCCTGTGGTGATTACGGGTANAGCTGTACAGAAATTAAGAGATGACCA	1002	
QY	970	agaagaagaagaagaagaagagtgtaagttgcattgggggaattgaaaaaaatggggaaaaat	1029	
Db	1003	AGAGAAGGAAGAAGGATTCGATAAGCTACTCAGAGAGAGAGCTTCGATGAGATAGGAATGGT	1062	
QY	1030	agtgctgtagtgctcgagtcgagttggggggtctctggcgaccctcggttgggatgtttcgtgac	1089	
Db	1063	GGTTTCATGGTGTGATCAATTACGGTTTTGAATCATAGATCGATAGGTTGTTTCGTGAC	1122	
QY	1090	gcatgtgggtggaaactcgctgtggagagacttgagttcgcggaattccgggtgggtggcggt	1149	
Db	1123	GCATTCCGGTGGAACTCTACGCTGGAGAGCTTGGTTTCAGAGATTCGCGTGGTGGCGGTT	1182	
QY	1150	gccgcagtggtttgatcatcagacgacgaaatgcgaagcctgattgaggaatcgctggggggaagg	1209	
Db	1183	TCGCGAATGGAATGATCAGATGACGACGCGAGCTTTAGAGATTGTTGGAACAACAGG	1242	
QY	1210	ggtgagagt-----gagaatgaatgaaggggtgggggttgatgtgatgtgagatagaagaag	1263	
Db	1243	TCTAAGAGTGATGGAGAAGAAGGAAGAAGAGAGTGTGGTGGTGGATCTAGAGAGAT	1302	

Qy	1264	gtgtgtgagatggtgtagtgatgggtgggtgcagaagaccacaaactagttagagaaatgccat	1323
Db	1303	ACGGCGGTGCATTGTAGGAGAGTATTGTAGAGACAAGCGGAGGAGTTTGTAGAGAAATGCCAC	1362
Qy	1324	caaatggaagactttggccagacagccatggatagatggatggtatcttcaactcaacaatct	1383
Db	1363	GAGGTGGAAGGATTAGCGCGGAGGCTGTGAGAGAGGAGGCTCTCTCTTTAATCACT	1422
Qy	1384	caacgcctttct	1395
Db	1423	CAAAAGCTTTGT	1434
RESULT	14		
LOCUS	A62523	1513 bp	DNA
DEFINITION	Sequence 28 from Patent WO9716559.	PAT	09-MAR-1998
ACCESSION	A62523		
VERSION	A62523.1	GI:3716430	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 1513)		
AUTHORS	Van, A.K., Marillia E., Peferoen, M., Grootwassink, J.W., Reed, D.W., Hemmingsen, S.M., Kolenovsky, A.D., Underhill, E.W. and Macpherson, J.M.		
TITLE	Plants with reduced glucosinolate content		
JOURNAL	PATENT: WO 9716559-A 28 09-MAY-1997;		
FEATURES	PLANT GENETIC SYSTEMS NV (BE)		
SOURCE	Location/Qualifiers		
	1..1513		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
	69..1469		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAA03659.1"		
	/db_xref="GI:3716431"		
	/translation="MAETVTITATNSKGHVVLVPVPGVHLPWVQFAKRLVSKGVKVI VTIATTTASSISTPVSVPVETISDGHDFLLIGVPGVSIDATSESLNGSETLTRYI SKFKSTSPIDSLVYDSFLPKGLEVARNSISAAAFPTNLTVCVSLKPFVSGEPFLP ADPASAPYLVRGLPALSYDELPSFVGRHSSHAHGRVLLAQFRNHEDADMLFVNSFE GLETQGEVGESEAMRATLIGPMIPSAIDGRKDDKQGGSLMKLPSEECMEWLDTR LSKSVYVSGSFQILFPEKGLAEVAKALQESNFWLWIKHAHTAKLPEGEVEATKDR ALLVSWNCGFLVLAHVSIGCTFLTCGNWSTLEGLSLGVPMVGVQWSDQMDKAFVEE VWRGYTRAKEAGGGVGVYDVVYRCLRGVMEGESSVEIRESSKKWKDLAVKAMSEGG SDRSTNEFVESLGGKH"		
BASE COUNT	368 a	356 c	399 g
ORIGIN	390 t		
Query Match	9.7%	Score 142.4;	DB 5; Length 1513;
Best local similarity	47.2%	Pred. No. 3.7e-21;	
Matches	634; Conservative	0; Mismatches 666;	Indels 42; Gaps 5;
Qy	39	gcgcgcgtgctgtagcaacgctttccgcgcgcgaagccacataaatccgcctcccaattcg	98
Db	112	GCACGTCGTGGTCTTACCTTACCAGTCCAAGGCCACCTCAACCAATGGTCCAGTTCC	171
Qy	99	ccaagagactcctataaagccgcgcgcactgactgactgcttttttcagagcgtttatgatggc	158
Db	172	CTAAAGCGCTAGTCTCCAAAGCGGTCAAAAGTCACAAATCGCCACACCTACACCGGCT	231
Qy	159	gcgcgcgtgcccacaacagactccgcgcgtgcgcggaaacccacccgcctcgactctgctg	218
Db	232	CTCCATCTCCACTCCCTCCGTCCTCGTTCGAACCAATCTCGAGGCCACGACTTCATCC	291
Qy	219	cgtctctcgaagcgtctacgaacagcgggctgaagcccgccgcgcggagcgtctacatgt	278
Db	292	TCAATAGCGCTCCCGCGCGTCAAGCATTCGACATCGACGATCTCCGAATCTTTCAAGCTCAACGGCT	351
Qy	279	ccagaatgaagccgcgcgcgtctccagagcctttaagaacacctctcttctcaacacgcacgc	338

```

Db 1399 CGATCAGTGAAGGAGGAGCTC 1420
RESULT 15
A72396 1513 bp DNA PAT 15-OCT-1999
LOCUS Sequence 28 from Patent EP0771878.
DEFINITION A72396
ACCESSION A72396
VERSION A72396.1 GI:6063728
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1513)
AUTHORS Van, A. K. and Peferoen, M.
TITLE PLANTS WITH REDUCED GLUCOSINOLATE CONTENT
JOURNAL Patent: EP 0771878-A 07-MAY-1997;
PLANT GENETIC SYSTEMS NV; CANADA NAT RES COUNCIL
FEATURES
source Location/Qualifiers
1..1513
/organism="unidentified"
/db_xref="taxon:32644"
69..1469
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB58422.1"
/db_xref="GI:6063729"
/translating="MAETTTTATNSKHVVVLPYPVQGHNLPMVQAKRLVSKGVK
VTATTTTASSISTPSVSEVPEISDHPDILIGVPGSIDAYSESPKLNSETLTVI
SKFKSDSPIDSLVYDFLPWGLVARNSISAAAFNTNLTGSLVKRPFVSGFPLP
ADPASPVLVRLGALPSYDELPSFVRSHSSHAHGRVLLNFRHEDADWLFWNSPE
GLETQCEVGESEAMBATLIGPMIPSNVLDGRIKDKGYSGLAKPLSESCMEWLDTK
LSKSVFVSFGSGLFELQALAEAKALOESNPNFLWIKRAHIAKLPEGFVATDKR
ALLVSWCNLEVLHVSIGCFILTHGWMNSTLEGLSGVPMVGPQNSQDMNDKAFVEE
VMRVGRAKEEGAGGVYKSDVVVRLGRVMEGSESVIRESKSKKKDKLAVKAMSEGS
SDRSINFVESLGKKH"
BASE COUNT 368 a 356 c 399 g 390 t
ORIGIN
Query Match 9.78; Score 142.4; DB 5; Length 1513;
Best Local Similarity 47.28; Pred. No. 3.7e-21;
Matches 634; Conservative 0; Mismatches 666; Indels 42; Gaps 5;
QY 39 gccggtgctgtcgtagcaagtttccggcgcaagccacacataaatcccgctcccaatcgg 98
Db 112 GCCAGGTGCTGCTTACCTTACCAGTCCAAGGCCACCTCAACCCCAATGGTCCAGTTCG 171
QY 99 ccaagagactcctaaagccgagactgaactcgttttttttttttttttttttttttttttt 158
Db 172 CTAAAGCCTAGTCTCCNAAGGCGTCAAGTCAATCGCCACACCACTACACGGCT 231
QY 159 gccgcatggccaaacacacagctcccgctcgccggaacacacacacacacacacacacac 218
Db 232 CCTCATCTCCACTCCCTCCCTCCGTCGTCGAACCAATCTCCGACGCCACGACTTCATCC 291
QY 219 cgtttccgaggttaagcaagcggggtgaagcccgcgcgcgcgcgcgcgcgcgcgcgcgcg 278
Db 292 TCATAGGCGTCCCGGCGCTCAGCATCGACGATCTCCGATCTTCAGAGCTCAACGGCT 351
QY 279 ccgagatgaagcccgcgctcccgagccttaagaacacacacacacacacacacacacacac 338
Db 352 CGGAACCTTAACCCGAGTATCTCAAAATTCAAATCCACAGATTCACCCATCGA---- 406
QY 339 tcaatttgctgtactaccacaccttttgcattgagcggcgcgcgcgcgcgcgcgcgcgcgcg 398
Db 407 ----TTCATTAGTACGACTCTTTCTCCCGTGGGACTCGAAGTTCGGAGATCTAACT 462
QY 399 agtctccgacgccttctctggttcgagcccgccacgctgctgt----gcataaccact 455
Db 463 CCATCTCAGCTGCTGCTTCTTCTACCAACAACCTCACCCTGCTGCTGCTTCTACGCAAT 522
QY 456 tctacttcaagcgctagcagcagagatgcagcccggttcccaattcgaattcagctccctc 515

```

Db 523 TCGTCTCCGGTGAGTTCTCTCCCGCGTGATCCCGGTTCCGGCGCGGTATCTCGTCCGTG 582
QY 516 ggcttccatccctgagcagcagcttctccgacgtttctgctgctgcgacgcgagaga 575
Db 583 GCTTACCGGCTTTGAGCTACGACGAGCTTCTCTCTCGTCCGACGCTCACTCGTCGAGCC 642
QY 576 gattccggttgatgataagagagaagcttttagacggtgaagagaagagcgaag 635
Db 643 ACCGGAACACCGGAGAGTCTTCTGAACAGTTCCTAACACACGAGATGCTGATTGCC 702
QY 636 tatgtgaaacacgttttgatcggttgagcccgatgcactcacggctattgatatgattg 695
Db 703 TGTTCGTCAACAGCTTCGAAGGTTAGAGACACAAAGTTGTGAAGTTGGAGATCAGAGG 762
QY 696 agtctgacg-----ggatcgggcggttgattccctccgcttcttggacgcggaagatc 749
Db 763 CGATGAGCGGACGTTGATCGGACCTATGATACCATCTGCTTATCTCGAGCGCCGAATCA 822
QY 750 cctccgaaacgtcttaccggcgcgatcttttcgaaaaatcggagagagaataactcgctg 809
Db 823 AAGACGATAAAGGCTACGGCTCGAGCCTGAT---GAAGCGGCTCTCGGAGGAGTGATGG 879
QY 810 agtgggtgaactcgaagccgaatacttcggtggtgctgctgcttggagcggttttga 869
Db 880 AGTGGTTAGACACTAAGCTGACCAAGTCGGTGGTTTTGTTGTTGCTTCTTGGGA 939
QY 870 ggtttccaaagcacaatgaagagattgggaaaggctattagcctgcggaaagccct 929
Db 940 TCCTCTTTTGAGAGCAACTCGTGGTGGCAAGGGGTTACAAAGATCCAACTTTAACT 999
QY 930 ttctatggatgatacgaagacagaagaatgacgacgagcgaagagaagaagaagaag 989
Db 1000 TCTTGTGGGTGATCAAGANAGCTCATATAGCGAAGTTACCAGAAGGTTTGTGGAAGCTA 1059
QY 990 agttggttgcatggggaattgaaaaaatgggaaaaatagtgtcgtggtgctgcagct 1049
Db 1060 CC-----AAAGACAGAGCGTTGCTGTGTTTCTTGTGTAAACAGC 1098
QY 1050 tggaggttctgcgaccctcgctgggatgtttcgtgacgcatttgggtggaactcgg 1109
Db 1099 TTCAGGCTTTAGCTCATGATCGATAGGTTGCTTTTGTACTCACTCGGTTGGAACTCGA 1158
QY 1110 ctgtgagagcttgagttgcgggattccggtggtggtggtgcgcagtggttggatcaga 1169
Db 1159 CGTGGAGAGGATTGAGTTGGAGTTCCGATGGTGTGCTGCTCAGTGGAGTGATCAGA 1218
QY 1170 cgaacgaatgcgaagctgattgaggtggtggtggtggtggtggtggtggtggtggtggt 1229
Db 1219 TGAATCATCTAAGTTTGTGGAGAGGTTTGGAGAGTTGGGATGAGGATAGAGCGAAG 1278
QY 1230 aaggggggtgggttgatggatgagatgagaagaggtgtgtgtgagatggtgatggatg 1289
Db 1279 CTGGGGGAGGAGTTGTGAAGAGCCGATGAGGTGGTGGTGTGTTGTGAGAGGAGTGTGGAAG 1338
QY 1290 gtgacaagaaccaaactagttagagaaaaatgcatcaaatggaagactttggccagacaag 1349
Db 1339 GAGAGAGTAGTGTGAGATTAGAGAGAGTCTTAAGAGAGTGGAAAGATTGGCTGTGAAGG 1398
QY 1350 ccattggatgattgattcttc 1371
Db 1399 CGATGAGTGAAGGGAAGCTC 1420

Search completed: August 1, 2000, 12:50:46
Job time: 54908 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:16:53 ; Search time 236.66 Seconds
(without alignments)
1558.282 Million cell updates/sec

Title: US-09-147-955-3
Perfect score: 1474
Sequence: 1 accaaaccacaaac

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			Description	
	Score	Match	Length	ID	Description
1	1474	100.0	1474	1	X02827
2	1327.8	90.1	1506	1	X02826
3	680	46.1	2062	1	X02828
4	446.8	30.3	1671	1	X02829
5	439.8	29.8	2105	1	X02831
6	142.4	9.7	1513	1	T66166
7	121.6	8.2	1731	1	T10106
8	116.8	7.9	566	1	T66173
9	109.4	7.4	568	1	T66174
10	75.8	5.1	432	1	T68693
11	70	4.7	1624	1	T17054
12	70	4.7	1626	1	X24873
13	66.4	4.5	1597	1	T23108
14	63.8	4.3	940	1	T66178
15	58.6	4.0	1650	1	T74684
16	57.2	3.9	1578	1	T62123
17	54.2	3.7	794	1	T66179
18	51.2	3.5	1738	1	Q58241
19	49.8	3.4	1607	1	V49609
20	48.6	3.3	405	1	T23852
21	48	3.3	28598	1	T06769
22	48	3.3	28958	1	T89956
23	48	3.3	49377	1	V05287
24	47.6	3.2	2064	1	Q52638
25	46.2	3.1	390	1	Q21833
26	46.2	3.1	390	1	Q36859
27	44.2	3.0	7193	1	V50431
28	43.8	3.0	13987	1	T80415
29	43.8	3.0	44377	1	T78508
30	43.8	3.0	44377	1	T80414
31	43.6	3.0	1483	1	X27947
32	43.6	3.0	1493	1	X37957
33	43.6	3.0	2871	1	T85076

ALIGNMENTS

RESULT	1	
XO2827	XO2827 standard; cDNA; 1474 BP.	
AC	XO2827;	
DT	14-MAY-1999 (first entry)	
DE	W90905287 Seq ID 2.	
KW	Plant: flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.	
OS	Perilla frutescens.	
FT	Key	
FT	Location/Qualifiers	
FT	CDS	
FT	29..1360	
FT	/*tag= a	
FT	/product= "protein with flavonoid 5-transglycosylation activity"	
PN	W90905287-A1.	
PD	04-FEB-1999.	
PF	16-JUL-1998; J03199.	
PR	25-JUL-1997; Jp-200571.	
PA	(SUNR) SUNTORY LTD.	
PI	Gong; Z, Mizutani M, Saito K, Tanaka Y,	
PI	Yamazaki M;	
DR	WPI: 99-142940/12.	
DR	P-PSDB; W92948.	
PT	Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be	
PT	used to transform plants for improvement of plant coloration	
PS	Disclosure; Page 53-56; 89pp; Japanese.	
CC	This invention describes the isolation of plant proteins which have	
CC	flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used	
CC	in the production of plant varieties with improved colour, for example in	
CC	the production of flower varieties with desired coloration and cut	
CC	flowers harvested from them.	
CC	Sequence 1474 BP; 369 A; 348 C; 438 G; 319 T;	
Sequence	1474 BP; 369 A; 348 C; 438 G; 319 T;	
Query Match	100.0%; Score 1474; DB 1; Length 1474;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1474; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 accaaacacaaacaaatttcacaaaaattgtccgcgcgcgtctctagcaagtt 60	
Db	1 ACCAAACCAAAACAAAATTTCCACAAAATGGTCCGCCGCCGTCTGCTAGCAAGTT 60	
Qy	61 tcgcgcgcgcgcgcgcacataatccgcgcctccaattcgcaagagactcctaaagccgg 120	
Db	61 TCGCGCGCAAGGCCACATAAATCCGCCCTCCCAATTCGCCAAGAGACTCCTAAAGCCGG 120	
Qy	121 cactgacgtcactgttttttcacagcgcgttatgatggcgcgcgcgcacacacgcctc 180	
Db	121 CACTGACGTCACGTTTTTTTTCAGGACGCGTTTATGATGGCGCGCGATGCGCAACACGCGTC 180	
Qy	181 cgcgcgcgcgcgcgcgcacccgcgcctcgcactcgtggcgattctccgcgcgcgcgcgcgc 240	
Db	181 CGCGCGTCGCGGAAACCCACCGGGCCTCGACTGCTGGTGGCGCTTCTCCGACGGCTACGACGA 240	
Qy	241 cgggctgaagccgc 300	
Db	241 CGGGCTGAAGCCGCGCGACGGAAGCGCTACATCTCCGACATGAAGACCGCGCGCTC 300	

```
QY 301 cgaggccttaagaaacotctctcacaacagcagcgtcactttcgtcgtactccca 360
DB 301 CGAGGCGCTTAAGAAACCTCTCTCAACACACGACGACGTCACCTTCGTCGTACTCCCA 360
QY 361 cctcttgcatggcgggaggtggcggtttgttccacagtcaccagccgcctctctcg 420
DB 361 CCTCTTGTGATGGCGCGGAGGTGGCGGTTTGTCCACAGTCCCGACCGCCCTCTCTG 420
QY 421 ggtcgagccgcgcacccgtcgtgtgcatataaccactctacttcaacgctcagcaga 480
DB 421 GGTGAGCCGCCACCGTGTGTGTCATATACCACTTCTACTTCAACGGCTACGACAGCA 480
QY 481 gatcagcccggttcccaatgaattcagctccctcggtctccatccctggagcagcgag 540
DB 481 GATGAGCCGCTTCCATGAATTCAGTCCCTCGGCTTCCATCCCTGGAGCAGCGCAG 540
QY 541 tcttcagcgtttctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
DB 541 TCTTCCGACGTTTCTGCTGCTCGACCGCGGAGAGATTCCGGTTGATGATGAAGGAGA 600
QY 601 gctggaactttaacggtgaagagagcgaggaagcattggtgaacagctttgatcggt 660
DB 601 GCTGGAACCTTACACGGTGAAGAGAGCGGAAGTATTGGTGAACAGCTTTGATGCGTT 660
QY 661 ggagcccgatgcactcgcgtattgataggtatgaggtatcggtatcggtcggtgat 720
DB 661 GGAGCCCGGACGCTACCGCTATTGATAGTATGATGATGATGATGATGATGATGATGAT 720
QY 721 tccctcgcgcctcttgacggcgaagatccctccgaaacgtcttcacggcgagctcttt 780
DB 721 TCCCTCGCCCTTCTTGACGCGGAAGATCCTCCGAACGCTTACGCGCGGATCTTTT 780
QY 781 cgaataatcgagagagataacgcgtgagtggttgaaactcgaactgaacgcgaatactcgg 840
DB 781 CGAAATTCGGAGGAGAAATACGCGTGGAGTGTGTGAACCTGAACGCGGAATCTTCGGT 840
QY 841 ggtgtatgtcgtttggagcgcttttgaggtttccaaaggcacaataaggagagattgg 900
DB 841 GGTGTATGTCTGCTGGGAGCGTTTGTAGGTTTCCAAAGGCACAAATGGAAGAGATTGG 900
QY 901 gaaaggcattatagcctcggaaggcccttttatgatatgatacagagaagaatga 960
DB 901 GAAGGGCTATTACCTCGGAAGGCGCTTTTATGGATGATACGAGAACAGAAATGA 960
QY 961 cgaaggcagaagaagaagaagaggttgagttgattgggggaattgaataaata 1020
DB 961 CGACGGCGAAGAAGAAGAAGAAGAGATTGACTTGCATTGGGGAATTGAAAAAAT 1020
QY 1021 ggggaataatgctcgtggtcgtcgagttgaggttctggcgaccctggttgggatg 1080
DB 1021 GGGGAAATAGTGTGTTGCTGCGAGTTGGAGGTTCTGGCGCACCTCGCTTGGGATG 1080
QY 1081 ttctgtacgcgaattgtgggtggaactcggctgtggagagcttgagttcgggattccggt 1140
DB 1081 TTTCTGTAGCGCATTTGCTGGTGGAACTCGGCTGTGGAGAGCTTGAGTTCGGGATTCGGT 1140
QY 1141 ggtggcgtgcccagtggtttgatcagacgacgaatgcgaagctgattgagatgcgtg 1200
DB 1141 GGTGCGGTGCGGAGTGGTTGATCAGACGAGCAATGGAAGCTGATTTGAGGATGCGGTG 1200
QY 1201 ggggacaggggtgagatggaatgaatgaagggtgggttggatgtgagataga 1260
DB 1201 GGGGACAGGGGTGAGATGAGATGAATGAATGAAGGGGTGGGTTGATGATGTGAGATAGA 1260
QY 1261 aaggtgtgtgagatggtatgggttgaaagacccaactagtgagagaataatgc 1320
DB 1261 AAGGTGTGTGAGATGATGATGATGGGTGACAAAGACCAACCTAGTGAGAGAAATGC 1320
QY 1321 catcaaatggagactttggccagacaaagccatgggatatgagttgattcttcaactcaaa 1380
DB 1321 CATCAATGGAAGACTTTGGCCAGACAAAGCCATGGGATAGGATGGATCTTCACTCAACA 1380
```

```
QY 1381 tctcaagcctttctcgtcagttgcacacttttbaatcgtcacaacagcggttcaaat 1440
DB 1381 TCTCAAGCCTTTTCTCGTCAAGTTGCACACTTTTAACTGCTCAACACGCGTTCAAT 1440
QY 1441 aaatatccctctccacttataaaaaa 1474
DB 1441 AAATATCCCTTCCACTTAAAAA 1474

RESULT 2
X02826
ID X02826 standard; cDNA; 1506 BP.
AC X02826;
DT 14-MAY-1999 (first entry)
DE W09505287 Seg ID 1.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Perilla frutescens.
FH Key
CDS Location/Qualifiers
FT /*tag= a
   /product= "protein with flavonoid 5-transglycosylation
   activity"
FT W09505287-A1.
PN 04-FEB-1999.
PD 16-JUL-1998; J03199.
PF 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
   Yamazaki M;
PI WPI: 99-142940/12.
DR P-PSDB; W92947.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
   used to transform plants for improvement of plant coloration
PS Disclosure; Page 49-52; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
   flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
   in the production of plant varieties with improved colour, for example in
   the production of flower varieties with desired coloration and cut
   flowers harvested from them.
CC Sequence 1506 BP; 364 A; 351 C; 449 G; 342 T;
SQ
```

```
Query Match 90.1%; Score 1327.8; DB 1; Length 1506;
Best Local Similarity 95.3%; Pred. No. 0;
Matches 1402; Conservative 0; Mismatches 37; Indels 32; Gaps 2;
```

```
QY 14 aaaaatttcacaaaaaggtccgcgcgcgtgctgtagcaacgtttccggcgcaagc 73
DB 2 AAAATTTCACAAAAATGGTCCGCGCGCGTCTGCTAGCAACGTTTCTTCGGCAGGC 61
QY 74 cacataaatccgcgcctcccaattccgcaagagactcctaaaagccggcactgaagcag 133
DB 62 CACATAAATCCGCGCCTCCAAATTCGCAAGAGACTCTTAAAGCCGGCACTGACGTCACA 121
QY 134 ttttcacagcgtttatgcatgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 193
DB 122 TTTTTCAGAGCGGTTTATGTCATGGCGCGCATGGCCGATGCCAACAACAGCCTCCGCGCTGCGGGA 181
QY 194 aacccacccggcctcgcacttctggttcctccgagcggctcgaacagcagcgggtgaagccc 253
DB 182 AACCCACCGCGCCTCGACTTCGTGGCGTTCTCCGACGGCTACGACGCGGTGAAGCCC 241
QY 254 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 313
DB 242 TCGCGCGAGCGGAGCGGCTACATGTCGAGATGAAGCCCGCGGCTCCGAGCGCTTAAGA 301
QY 314 aacctctctcacaacagcagcgttcactttctgctactctccacctctttgcatgg 373
DB 302 AACCTCTCTTCAACAACACGAGGCTCACGTTTCTGCTCTACTCTCCACGCTCTTTTCATGG 361
QY 374 gcggcgaggggtggcggtttgtcccaagtcgccagccgctctcttgggtcgagccgcgc 433
DB 362 GCGCGGAGGTGGCGGTGAGTCCCAAGTCCCGAGCGCCTTCTCTGGGTGCGAGCCGCGC 421
```

```
QY 434 accgtctgtgatataccactctacttcaacggtcagcagacagatcgagccggt 493
Db 422 accgtctgtgatatacttacttcttacttcaacggtcagcagacagatcgagccggt 481
QY 494 tccaatgaattcagctccctcgtctccatccctggagcagcagctctccgagcttt 553
Db 482 TCGGAGAAATTCAGTCCCTCGGCTCCACCCCTGGGAGCAGCAGCTTCGACCTTT 541
QY 554 ctgctccctgcagcccgagagatcccggttgatgtatgaagagagatcggaacttta 613
Db 542 CTGCTCCGAGACACCGAGAGATTCGCGTTGATGTATGAAGGAGAGCTGGAATTTA 601
QY 614 gacggtgaagagaagcgaagatattgtgaacacgtttgatcggttgagcccgatga 673
Db 602 GACGCTGAAGAGAGCGGAAGTGTGTGTGAACACGTTTGTGCTGGAGCCGATGCA 661
QY 674 ctacggtctatgatagtgatagtgatcggtatcggtatcggttgcctccgcttc 733
Db 662 CTCACGCTATTTGATAGGTATGATGTGATCGGGATCGGGCGTTGATTCCTCCGCTTC 721
QY 734 ttgagcggcgaagatccctccgaacgtcttcacgcgcgatcttttcgaaaaatcgag 793
Db 722 TTGAGCGCGGAGATCCCTCCGAACGCTTTACGGCGCGATCTTTTGGAAATCGGAG 781
QY 794 gaaataactgcgtgagtggttgtaactcgaagcgaacatcttcggttggtatgtcg 853
Db 782 GAGAAATACCTGCTGGAGTGTGTGACACGAAGCCGAATCTTCGTTGTTATGTGTCG 841
QY 854 ttgggagcgctttgaggtttccaaagcacaatggaagagattgggaagggctatca 913
Db 842 TTTGGAGCGCTTTTGGAGTTTCCAAAGCCACAAATGGAAGAGATTGGGAAGGCTATTA 901
QY 914 gctcggaagagccctttttatggtatgatacgaagaagaagaatgacagcggcgaaga 973
Db 902 GCTCTCGGAAGCGCGTTTATGATGATATGAGAACAGAGAAATGACGACGGC----- 955
QY 974 gaagaagaagaagaagttgagttgcatctggggaattggaataaaggggaataagtg 1033
Db 956 GAAGAAGAAGAAGAAGTGTGATTTGCTGGAATTTGGAATAATGGGGAATAATGTT 1015
QY 1034 tctggtgtcgcagtgaggttctgcgcacccctgcgttgggagttcttcgtgaagcat 1093
Db 1016 TCGTGTGCTCGCAGTGTGAGGTTCTGCGCACCCCTCGCTGGGATGTTTTCGTGACGCAT 1075
QY 1094 tgggttggaactcgcgtgtgagagcttgagttgcggatctccggtgggtggcggtccg 1153
Db 1076 TGTGGTGGAACTCGGCTCTGAGAGCTTGAGTTGCGGGGTTCCGGTGGTGGCGGTCCG 1135
QY 1154 cagtggttgatcagacagcaatcgaagctgattgaggtcggtgggggacaggggtg 1213
Db 1136 CAGTGTGTGATCAGACGACGAATCGAAGCTGATGAGGATGCTGGGGGACAGGGGTG 1195
QY 1214 agagtgagaatgaatgaaggggtgggttgatgagtgagatagagaatggtgtggag 1273
Db 1196 AGAGTCAGAATGAATGAAGGGGTGGGCTTCATGATCTGAGATAGAGAGTGTGTGGAG 1255
QY 1274 atggtgatgagtggggggtgacagacaaactagtgagagaaaaatgccatcaaatggaag 1333
Db 1256 ATGCTGATGGATGGGGGTGAGAAGAGCAAACTGCTGAGAAAAATGCCATAAATGGAAG 1315
QY 1334 actttggcagacaagcatggatagatgagttctcaactcaacaatctcaacgcttt 1393
Db 1316 ACTTTGGCCAGAGAGCCATGGGAGAGATGGATCTTCACTCAAGAATCTCAACGCTTT 1375
QY 1394 cttcgtcaagttgca-----cactttaatctgctcaaa 1427
Db 1376 CTTTCATCAAGTTGACGTGCTTTAATACAAAAATGGCTTTTCCACTTTTAACTTACTCAAA 1435
QY 1428 cagcgttcaataaataatccccctccactt 1458
Db 1436 CACCGGTTCAATAATAATATCCCTCTCACTT 1466
```

```
RESULT 3
X02828 standard; cDNA; 2062 BP.
ID X02828;
AC X02828;
DE 14-MAY-1999 (first entry)
DR WO905287 Seq ID 3.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Verbena hybrida.
FH Key.
FT CDS
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity".
FT
FT WO905287-A1.
PD 04-FEB-1999.
PP 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
PI WPI: 99-142940/12.
PI P-PSDB: W92949.
DR
DR Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
DR used to transform plants for improvement of plant coloration
PT
PT Disclosure: Page 56-60; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
CC
CC Sequence 2062 BP; 608 A; 354 C; 508 G; 592 T;
QY
Query Match 46.1%; Score 680; DB 1; Length 2062;
Best Local Similarity 70.5%; Pred. No. 4.7e-172;
Matches 1019; Conservative 0; Mismatches 370; Indels 57; Gaps 6;
QY 12 acaaaatttcacaaaaatggtccgcgcgctgctgtcagcaacgtttccgcgcaag 71
Db 9 AAAAAATAAAAAAATAATGACGAGCTCAGTCTCTTGCCACATTCACACACAGG 68
QY 72 gcaacataatcccgccctccaatccgaagagactctctaaagccggagcagctca 131
Db 69 GACACATAAATCCCGCCTTCAATTCGCCAAGCGTCTCGCAATGCCGACATTCAAGTCA 128
QY 132 cgttttcacagcagcttattgatgagcgcgcgcatggccaacacagctcccgcgctcg 191
Db 129 CATCTTCACACAGGCTACGCGATGCGCGCGCATGTCAGAAACCGCGCTGAGCTCAAAG 188
QY 192 gaaacccacccggcctcgacttcgtggttcctccgagcgtacgagcagggctgaagc 251
Db 189 G-----GCTCATCAATTTTGTGCTGTTTTCGACGGGTATGACGACGGGTACAGC 239
QY 252 ccgcgcgagcaggaagcagctacatgctcgagatgaagcccgcgctccgagccttaa 311
Db 240 CCGGAGACGATGGGAAGAACTACATGTCGGAGATGAAAGCAGAGGTATAAAGACCTTGA 299
QY 312 g-----aaacctctctcaacaacgacgagcgtcacttctcgctct 353
Db 300 GCGATACTCTTGACGCAATAATGTCGATCAAAAAAAGCAGCAAAATCAGTTCGTGCTGT 359
QY 354 actccacactctttgatggcgcgaggtggtggtggttccagctccacacgccc 413
Db 360 ACTCCCCACCTCTTTGCTGCGCGCGCAAGGTGGCGGTGAGTTCATCTCGGAGCGCGC 419
QY 414 ttctctgggtcgagcccgccacgctgctgtgtgcataaccactctacttcaacggcttacg 473
Db 420 TACTCTGGATTGAGCCAGCTACGGTGTGTGATATATTTTACTTTTATTTTCAACGCTATA 479
QY 474 cagacgagatcgagcgggttccaatgaaattca---gctccctcggttccatccctgg 530
Db 480 CGGACGAATTCGATGCGGGTTCGGATGCTATTTCATTGCCCGGAGGACTCCCGAGTGTGG 539
```



```

QY 829 gaaatcttcggtggtatgtcgtgttgggagcggtttgaggtttccaaaggccacaaat 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1165 AAATTCATCAGTGTGTTTATATATCTTTTGGGAGTCTAATGAATCCATCTATTATGCCAAAT 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 889 ggaagatgtggaaaggctattagcttcggaaggccctttttatgatgatacagaga 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 GGAGGAGATATCAAAAGSGTTGATAGACATAGGAAGCCGTTTTTATGGTGATANAAGA 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 949 acagaagaatgacagccgacgaagaagaagaagaagaagagtgagttgattgcattgggga 1008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1285 AAATCAAAAGGC-----AAAGAAGAAGAGATAAAAAGCTTGGTTGATTGAAGA 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1009 attgaaaaaaatgggaaaaatagtctgctggctcgagtcgagttcgaggttcggcgccacc 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1336 ATTGGAANAATAGAAAATAGTTCATGGTGGTTTCACAACTTGAAGTTCTAAAACATCC 1395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1069 tgcgttggagtggttcgcgcgcatgtggtgggaactcggctggtggagcttgagttg 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1396 ATCTTTAGGATGTTTCTTCTCATTGTGGATGGAATTCAGCCCTTAGAGAGTTTAGCTTG 1455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1129 cgggatcccggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1456 TGGAGTGCCACTTGTGCGCATTTCTTCANTGGACAGATCAAAATGACAAATGCCAAACAAAT 1515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1189 tgaggatgctggtgggggacaggggtgagagtgagagtgagagtgagagtgagagtgag 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1516 TGAAGATGTGTGGAAGAGTGGAGATAGAGTGAAGATGAAGATGCTGTGTGTAAG 1575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1249 atgtagatgagaagggtgtgtgagatggtgagtggtggtggtggtggtggtggtggt 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1576 TGAGGAATCAAAAGGTGTATTGAATGTTGATGGATGAGGAGAGAAAGGGGAAGAAAT 1635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1309 gagagaaatgcatcaaaatggaagacttggccagacgaagccatgggataggatggatc 1368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1636 GAGAAAGAAATCTAAGAAATGGAAGAAATTTGGCTAGAGAAAGCTGTGAAGGAAGTGGATC 1695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1369 ttcactcaacatctcaacgccttcttctgctcaagttgc 1407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1696 TTCACACAAGAATTTAAAGGCTTTTATTGATGATGTTGC 1734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
T66166
ID T66166 standard; cDNA to mRNA; 1513 BP.
AC T66166;
DT 15-JUL-1997 (first entry)
DE UDP-glucose:thiohydroximate S-glucosyltransferase cDNA clone pGL9.
KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
OS Brassica napus cv. Westar.
FH Key Location/Qualifiers
cds 69..1469
FT /*tag= a
FT 73
FT /*tag= b
FT /*note= "C is T in clone pGL6-14, causing an amino
FT acid change from Ala-2 to Val-2"
FT 98
FT /*tag= c
FT /*note= "G is A in clone pGL6-14; there is no amino
FT acid change"
FT 99..99
FT /*tag= d
FT /*note= "codon AAG (Lys) is present between clone
FT pGL9 bases 98 and 99 in clone pGL6-14"
FT 195
FT /*tag= e
FT /*note= "G is C in clones pGL6-14 and pGL3-22,
FT causing an amino acid change from Val-43
FT to Leu-43"
FT 245
FT /*tag= f
FT /*note= "C is G in clone pGL3-22; there is no amino

```

```

FT /*note= "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 292
FT /*tag= g
FT /*note= "T is C in clone pGL3-22, causing an amino
FT acid change from Leu-75 to Pro-75"
FT 323
FT /*tag= h
FT /*note= "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 331
FT /*tag= i
FT /*note= "A is G in clone pGL3-22, causing an amino
FT acid change from Glu-88 to Gly-88"
FT 345
FT /*tag= j
FT /*note= "A is C in clone pGL3-22, causing an amino
FT acid change from Asn-93 to His-93"
FT 354
FT /*tag= k
FT /*note= "G is C in clone pGL3-22, causing an amino
FT acid change from Glu-96 to Gln-96"
FT 368
FT /*tag= l
FT /*note= "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 377
FT /*tag= m
FT /*note= "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 398
FT /*tag= n
FT /*note= "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 410
FT /*tag= o
FT /*note= "A is T in clone pGL3-22; there is no amino
FT acid change"
FT 458
FT /*tag= p
FT /*note= "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 465
FT /*tag= q
FT /*note= "A is C in clone pGL3-22, causing an amino
FT acid change from Ile-133 to Leu-133"
FT 473
FT /*tag= r
FT /*note= "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 512
FT /*tag= s
FT /*note= "T is C in clone pGL3-22; there is no amino
FT acid change"
FT 515
FT /*tag= t
FT /*note= "A is T in clone pGL3-22; there is no amino
FT acid change"
FT 526
FT /*tag= u
FT /*note= "T is C in clone pGL3-22, causing an amino
FT acid change from Ala-153 to Val-153"
FT 568
FT /*tag= v
FT /*note= "C is T in clone pGL3-22, causing an amino
FT acid change from Pro-167 to Leu-167"
FT 587
FT /*tag= w
FT /*note= "A is G in clone pGL3-22; there is no amino
FT acid change"
FT 623
FT /*tag= x
FT /*note= "C is G in clone pGL3-22; there is no amino

```

```
FT conflict
FT /tag= y
FT /note= "G is C in clone pGL3-22; there is no amino
FT acid change"
FT 678..679
FT /tag= z
FT /note= "bases 678-679 are AT in clone pGL3-22,
FT causing an amino acid change from Arg-204
FT to Ile-204"
FT 685
FT /tag= aa
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 714
FT /tag= ab
FT /note= "A is G in clones pGL3-22 and pGL4-2,
FT causing an amino acid change from Ser-216
FT to Gly-216"
FT 759
FT /tag= ac
FT /note= "G is A in clone pGL3-22, causing an amino
FT acid change from Ala-232 to Thr-232"
FT 761
FT /tag= ad
FT /note= "G is C in clone pGL3-22, causing an amino
FT acid change from Ala-232 to Thr-232"
FT 766
FT /tag= ae
FT /note= "G is A in clones pGL3-22 and pGL4-2,
FT causing an amino acid change from
FT Arg-234 to Lys-234"
FT 794
FT /tag= af
FT /note= "A is C in clone pGL3-22; there is no amino
FT acid change"
FT 809
FT /tag= ag
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 814
FT /tag= ah
FT /note= "G is C in clone pGL3-22, causing an amino
FT acid change from Gly-249 to Ala-249"
FT 848
FT /tag= ai
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 901
FT /tag= aj
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 936
FT /tag= ak
FT /note= "G is A in clone pGL3-22, causing an amino
FT acid change from Gly-290 to Arg-290"
FT 959
FT /tag= al
FT /note= "C is A in clone pGL3-22; there is no amino
FT acid change"
FT 968
FT /tag= am
FT /note= "G is A in clone pGL3-22; there is no amino
FT acid change"
FT 973
FT /tag= an
FT /note= "A is C in clone pGL3-22, causing an amino
FT acid change from Lys-302 to Thr-302"
FT 1013
FT /tag= ao
FT /note= "C is T in clone pGL3-22; there is no amino
FT acid change"
FT 1117
```

```
FT conflict
FT /tag= ap
FT /note= "G is T in clones pGL2-7 and pGL2-25,
FT causing an amino acid change from
FT Val-350 to Glu-350"
FT 1121
FT /tag= aq
FT /note= "G is T in clone pGL2-7; there is no amino
FT acid change"
FT 1130
FT /tag= ar
FT /note= "C is T in clone pGL2-7; there is no amino
FT acid change"
FT 1161
FT /tag= as
FT /note= "T is C in clone pGL2-7; there is no amino
FT acid change"
FT 1169
FT /tag= at
FT /note= "A is G in clone pGL2-7; there is no amino
FT acid change"
FT 1193
FT /tag= au
FT /note= "T is G in clone pGL2-7; there is no amino
FT acid change"
FT 1202
FT conflict

Query Match 9.7%; Score 142.4; DB 1; Length 1513;
Best Local Similarity 47.2%; Pred. No. 1.2e-28;
Matches 634; Conservative 0; Mismatches 666; Indels 42; Gaps 5;

QY 39 gcgcgtgtgttagcaacgttttcgcgcgaagccacataatccgcgcctcaattgc 98
DB 112 GCCACGTGTTGTTTACCTTACCAGTCCAAAGCCCTCAACCCAAATGTCAGTTCG 171
QY 99 ccaagagactcctaaagcgcgcactgcgtcactcgtttttccacagagcgtttatgc 158
DB 172 CTAACGCCCTAGTCTCCAAAGCGTCAAAAGTCAACATCGCCACCACCTACACGCC 231
QY 159 gcgcgatggcacaacacagctcgcgcgtgcgcgaacccacgcgcgcctgcactgtg 218
DB 232 CTTCCATCTCCACTCCTCCTCGTCTCGTCAACCAATCTCCGACGGCCACGACTTC 291
QY 219 cgtctccagcggctacgcagcgggtgcgaagccgcgcgcgcgcgcgcgcgcgcgc 278
DB 292 TCATAGGCGTCCCGCGCTCAGCATCGAGCGATACCTCGGAATCTTCAAGCTCAACGC 351
QY 279 ccgagatgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 338
DB 352 CCGAACCCTTAACCCGAGTAATCTCAAAATTCAAATCCACAGATTCACCAATCGA 406
QY 339 tcactttcgtcgttactccaccttttgcagcgcgcgcgcgcgcgcgcgcgcgcgcgc 398
DB 407 ----TTCAATAGTCTACGACTCTTCTCCTCCGCTGGGACTCGAAGTCGAGATCTAA 462
QY 399 acgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 455
DB 463 CCATCTCAGCTGCTGCTTTCTTCCACCAACAACTCACCGTTTGTCTGTCTAGCAAA 522
QY 456 tctacttcaacggctacgcagcagatcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 515
DB 523 TCGTCTCCGCTGAGTTTCTCTCTCCCGCTGATCCCGCTTCGCGCCGCTATCTCGTC 582
QY 516 gggttcaccccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 575
DB 583 GCTTACCGGCTTTGAGCTACGACGAGCTTCTTCTTCTCGTCCGACGCTACCTCGT 642
QY 576 gattccggttgatgatgaaggagagcgtggaaacttttagacgcgtgaagaaagcga 635
DB 643 ACGGGAACACGGGAGAGTTCTTCTGAACCACTTCCGTAACCAAGAGTCTGATTGGC 702
QY 636 tattggtgaacgcttttgatcggttgagccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 695
DB 703 TGTCTGTCACACGCTTCGAAGGGTTAGAGACACAAGGTTGTGAAGTTGGAGATTCAG 762
```

QY 696 agttgatcg-----gagtcggcgcttattccctccctcttcttgaacggcgaaagatc 749
 DB 763 CGATGAGGCGACCTTGTATCGGACCTATGATACCATCTCTTATCTGACGCGCGGAATCA 822
 QY 750 cctccgaacgtcttaccggtcgagctcttttgcgaatacggagagataactgcgtgg 809
 DB 823 AAGACGATAAAGGCTACGGCTCGAGCTGAT---GAAGCGGCTCTCGAGAGGTATGG 879
 QY 810 agtggttgaactcgaacgcaaaatcttcggtggtgtatgtcgtttgggagcgttttga 869
 DB 880 AGTGCTTAGACACTAAAGCTGAGCAAGTCGGTGGTGTGTTTCGTTGGTTCCTTTGGGA 939
 QY 870 ggtttccaaagcgcaaatggaagagatgggaagggggtattaccgcgcggaaggccct 929
 DB 940 TCCCTTTTGAGAAGCAACTCGTGAGGTGGCAAGGCGTTACAAGATCCAACATTAACT 999
 QY 930 ttttatgagatgacgagacagaaagaaatgacgacgcgcaagaaagaaagaaagaaag 989
 DB 1000 TCTTGTGGGTGATCAAGAAAGCTCATATAGCGAAGTTACCAAGAGGGTTTGTGGAAGCTA 1059
 QY 990 agttgagttgctgggggaattgaaaaaaatggggaataatagtcgtggtgctcgcaat 1049
 DB 1060 CC-----AAAGACAGAGCGTGTCTTCTTTGGTGTAAACGAGC 1098
 QY 1050 tggaggtcttgccgacccctcggtgggatgttctcgacgcatgtggtggaactcgg 1109
 DB 1099 TTGAGGTTTAAAGCTATGATATGATAGTATGCTTTTGACTCAGTGGGTTGGAACTCGA 1158
 QY 1110 ctgtggagagcttgagttgcggatccggtggtggtgcgcgcagtggtttgatcaga 1169
 DB 1159 CGTTGGAAGGATTGAGTTTGGAGATCCGATGTTGCTGCTCAGTGGAGTATCAGA 1218
 QY 1170 cgaagaaatgcaagctgattgagatgcgtgggggacaggggtgagagtgagaatgat 1229
 DB 1219 TGAATGATGCTAAAGTTGTGGAGAGGTTTGGAGAGTTGGGTATAGACGGAAGGAGGA 1278
 QY 1230 aagggggtgggttgatggatgtgagatgaaagggtgtgtggagatggtgatggatggg 1289
 DB 1279 CTGGGAGGAGGATTGTACAGAGCGATCAGGTGTGAGGTGTTGAGAGGATGATGGAAG 1338
 QY 1290 gtgacaagaccactgtagagaaataatgcataatgggaagactttggccagacaag 1349
 DB 1339 GAGAGAGTAGTGTGGAGATTAGAGAGAGTTCTAAGAAAGTGAAGATTTGGCTGTGAAG 1398
 QY 1350 ccatgggatagatgatatcttc 1371
 DB 1399 CGATGAGTGAAGGAGGAAGCTC 1420

RESULT 7

T10106
 ID T10106 standard; cDNA to mRNA; 1731 BP.
 AC T10106;
 DT 27-OCT-1996 (first entry)
 DE Maize UDP-glucose indol-3-yl:acetyl:glucosyl transferase sequence.
 KW UDP-glucose-indol-3-yl:acetyl:glucosyltransferase; enzyme;
 KW transgenic plant; EC-2.4.1.121; crop improvement; corn; cereal;
 KW grass; IAGlu-transferase; ds.
 OS Zea mays.
 FH key Location/Qualifiers
 FT cds 57..1472
 FT /tag= a
 FT misc_binding 858
 FT /tag= b
 FT /note= "start of UDP binding region"
 PN W09600291-A1.
 PD 04-JAN-1996.
 PF 20-JUN-1995; U07820.
 PR 24-JUN-1994; US-265427.
 PA (RESE) RESEARCH CORP TECHNOLOGIES INC.
 PI Bandurski RS, Szczylowski K, Szareszen JB;
 DR WPI; 96-068875/07.

DR P-PSDB; R99999.
 PT UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein of
 PT - used in sense or anti-sense orientation for the genetic control of
 plant IAA levels and plant growth
 PS Claim 2; Page 21-23; 41pp; English.
 CC This nucleotide sequence encodes a UDP-glucose-indol-3-
 CC ylacetylglucosyltransferase (IAGlu-transferase) from maize (Zea mays).
 CC Over-expression of this sequence in transgenic plants will result
 CC in at least 1 of the following effects: inhibition of apical
 CC dominance; inhibition of stem elongation; inhibition of cell
 CC enlargement and increased numbers of stems per plant as compared
 CC with a wild-type plant. The ability to control the expression of
 CC this enzyme allows the control of free indoleacetic acid levels in
 CC plants, thereby effecting plant growth rates. Antisense constructs
 CC may be used to inhibit synthesis of this enzyme.
 SQ Sequence 1731 BP; 298 A; 528 C; 601 G; 304 T;

Query Match 8.2%; Score 121.6; DB 1; Length 1731;
 Best Local Similarity 52.8%; Pred. No. 4.5e-23;
 Matches 321; Conservative 0; Mismatches 26; Indels 18; Gaps 2;

QY 793 ggagataactcgtgagtggtgaactcgaagcgaatacttcggtggtgtagtgc 852
 DB 848 GGAGGATGCGTGCACCAAGTGGCTAGACACCAAGCCCGACCGCTCGTGGCCCTAGCTCTC 907
 QY 853 gtttgggagcgttttggaggtttccaaaggccaaatggagagattgggaagggtcatt 912
 DB 908 CTTTCGGCAGCCTCGCTCCCTGGCAACGCCAGAGAGGAGAGCTCGCGCGCGCTCCT 967
 QY 913 agcctcggaagcccttttattgagatgacgagaacagaagaatgacgacgagga 972
 DB 968 CGCCGCCCGCAAGCGTTTCCTGTGGGTGTGAGGGC-----CAGCAGCA 1012
 QY 973 agaagaagaagaagaagttgagttgagttgggaattgaaaaaaatgggaaaaatagt 1032
 DB 1013 GCACCAAGTCCCGGCTATCTCTGGCCGAGCGACGCGCGCGCGATGGTCT 1072
 QY 1033 gtcgtggtctcagattgaggtttctgacgacccctcggttgggtatttttcgtacgca 1092
 DB 1073 GCCTGTGTCCCGCAGCTGACGTCGTGGCCGACCCCGCGCTGGGCTGCTTCGTCAACCA 1132
 QY 1093 ttgtgggtggaactcggctgtgagagacttgagttcggggattccggtggtggcggtgcc 1152
 DB 1133 CTGCGGTTTGGAACTCCACGCTGGAGCGCTCAGCTTCGGCGTGCCTATGTGGCGATGGC 1192
 QY 1153 gcagtggtttgagcagcgaatgcgaagctgattgaggtcgttggggagacaggggt 1212
 DB 1193 GCTGTGGAGGACGACCGCAGCAACGCTCGGAACGTCGAGCTCGCTGGGGCGCGGGCT 1252
 QY 1213 gagatgagaatgaatgaagg---gggtggggttgatgagatgagatgagaagggtgct 1269
 DB 1253 CGCGCGCGCGCGATGCTCGCGCGCGCTGTTCTTCGCGGGGAAGTGGAGCGGTGGCT 1312
 QY 1270 gagatggtgagtgatgggggtgacagaacaaactagtgagagaaaatgcatcaaatg 1329
 DB 1313 CGCGCGCTCATGGACGGGGCGAGGGCGGCTGTGCTGCACCAAGCGCGGGGGAATG 1372
 QY 1330 gaagactttggccagacaagccatgggataggatgagattcttcaactcaacaactcaagc 1389
 DB 1373 GAGGACAGGGCTCGCGCGCGGTGGCACCCGCTGGCAGCTCTGACCGCAACCTGGACGA 1432
 QY 1390 cttttcttc 1397
 DB 1433 GTTCGTGC 1440

RESULT 8

T66173
 ID T66173 standard; cDNA to mRNA; 566 BP.
 AC T66173;
 DT 15-JUL-1997 (first entry)
 DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGL2-7.

PI Manning K;
 DR WPI: 97-332787/30.
 PT Vectors encoding new enzymes for regulating ripening of fruit - used
 PT particularly to improve storage properties, processing
 PT characteristics etc, especially in strawberries
 PS Claim 1: Page 23-24; 43pp; English.
 CC This cDNA sequence comprises a UDP-glucuronosyl transferase (BRT
 CC 1b) cDNA clone isolated from a ripe strawberry fruit cDNA library
 CC by differential screening. 9 Ripening-enhanced clones (see T68687-
 CC 95) clones were isolated from the library, and are deposited at
 CC NCIMB 40933. Any of these 9 sequences can be used as a regulatory
 CC sequence in claimed vectors for genetic transformation of plant
 CC cells to regulate fruit ripening. The vectors also comprise a
 CC promoter and transcription termination sequence. Also claimed are
 CC plants (especially strawberry) and propagating materials containing
 CC these vectors. Stable integration of the regulatory sequences, or
 CC their complements or genomic equivalents, into a plant is used to
 CC modulate fruit ripening by overexpression or downregulation of an
 CC endogenous plant gene. Slowing ripening of fruit improves
 CC resistance to damage during harvesting, packaging and transport.
 CC It also extends shelf life, improves storage properties, processing
 CC characteristics, flavour and aroma, and modifies colour and
 CC increases resistance to post-harvest fungi.
 SQ Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;

Query Match 5.1%; Score 75.8; DB 1; Length 432;
 Best Local Similarity 60.4%; Pred. No. 4.2e-11;
 Matches 125; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1015 aaaaatggggaataatgctgctggtgctgcagttggaggttctggcgccacctgcgtt 1074
 DB 105 AGACAGAGGCAAAAGTGTGCAATGGAGTCCACAGAGAGAGATTTGGAGCATCTTCGAC 164
 QY 1075 gggatgttcgtgacgcattgtgggtggaactcggctgtgagagcttgattgcgggat 1134
 DB 165 GCGTGTGTTGTGACTCATTTGCGGGTGGAACTCAACCATGGAGTCACCTCAGGAAT 224
 QY 1135 tccggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1194
 DB 225 GCCCGTGTGTCATCCCAATGGGTCACCAAGTGACCGAGCCGCAAGTATTGGTCGA 284
 QY 1195 tgcgtgggggacaggggtgagagtga 1221
 DB 285 CGAGTTAAAGTGGGAGTAAGAATGTG 311

RESULT 11
 V17054
 ID V17054 standard; cDNA; 1624 BP.
 AC V17054;
 DT 08-JUN-1998 (first entry)
 DE Glucosyl transferase (Gtase) encoding wound inducible gene (TW11).
 KW Glucosyl transferase; Gtase; TW11; tomato; signalling pathway;
 KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;
 KW plant defence protein; plant response; tobacco; rice; ss.
 OS Lycopersicon sp.
 FH Key Location/Qualifiers
 FT CDS 2..1410
 FT /*tag= a
 FT /product= "glucosyl transferase"
 FT /note= "encodes Gtase from amino acid position 5
 FT shown in W47172"

PN W09745546-A1.
 PD 04-DEC-1997.
 PF 30-MAY-1997; G01473.
 PR 31-MAY-1996; GB-011420.
 PA (Utrio-) UNIV YORK.
 PI Bowles DJ, Calvert CM, ODonnell PJ, Roberts MR;
 DR WPI: 98-032653/03.
 DR P-PSDB: W47172.
 PT Tomato wound inducible (Twil) gene encoding glucosyl transferase -
 PT useful to develop products that alter signalling pathways in plants

PT by altering of salicylic acid, jasmonic acid or ethylene
 PS Claim 1; Fig 1; 52pp; English.
 CC This wound inducible gene (TW11) isolated from wounded tomatoes encodes
 CC a glucosyl transferase (Gtase) protein. This TW11 gene can be used to
 CC identify homologue Gtase encoding genes isolated from tobacco and rice.
 CC A microbial host can be transfected or transformed with a vector
 CC containing the Gtase encoding nucleic acids. The products can be used to
 CC interfere with Gtase and therefore alter signalling pathways in plants,
 CC specifically tobacco, rice or tomato plants by altering levels of
 CC salicylic acid, jasmonic acid or ethylene. This can induce the
 CC production of plant defence proteins such as pathogenesis-related (PR)
 CC and proteinase inhibitor (PIN) proteins which regulate plant development
 CC (plant growth, reproduction and senescence) and improve plant response to
 CC pathogens.
 SQ Sequence 1624 BP; 545 A; 252 C; 372 G; 455 T;

Query Match 4.7%; Score 70; DB 1; Length 1624;
 Best Local Similarity 49.8%; Pred. No. 2.6e-09;
 Matches 221; Conservative 0; Mismatches 205; Indels 18; Gaps 1;

QY 791 gagagaataaactcgtcgagtggtggaactcgaagccgaatcttcggtggtgatgtg 850
 DB 779 GATGACACCGCTGCTTGAATGCTTGATTCGAGAAATCAAGTCCATTTGTTATGTT 838
 QY 851 tggtttgggagcgttttggaggtttccaaaggccaaatggaagagattgggaaaggccta 910
 DB 839 TGTGTTGGAAGTACAGCAGATTTTCACTACAGCAGACAGATCAAGAACTTTCCTATGGGGCTA 898
 QY 911 ttgacctcggaagcccttttctggtgatgatacagagaacagaagaatgacgacggcga 970
 DB 899 GAAGCCTCTGACAAAGATTTTCATTTGGTGTATCAGAACAGGCAATGAAGATTGGCTCCA 958
 QY 971 gaagaagaagaagaagaagattgagttgcattgggggaattgaaaaaaatggggaataa 1030
 DB 959 GAAGGATTCGAGGAAAGAAC-----AAAAAGAAAAGGTTTAAATCATA 1000
 QY 1031 gtgtcgtggtgctgcagttggaggttcttcggcgccacctgcgttgggatgtttcgtgacg 1090
 DB 1001 AGAGGATGGGCAACCCCAAGATGTGATTTCTTGATCAGAAAGCTATTGGAGCTTTTGTACT 1060
 QY 1091 cattgtgggtggaactcggctgtgagagcttgagttcggtgattccggtggtggtggtg 1150
 DB 1061 CATTTGATGGAATCTGACACTGGAGGAGGANTATCAGCAGGGGTACCAATGTTGACATGG 1120
 QY 1151 ccgcagtggtttgatcagacgacgaatgctgaagctgattgaggtgctgtgggggacaggg 1210
 DB 1121 CCAGTATTTCGGAACAGTTTTCAATGAGAAGCTTGGTGACTGAGTAATGAGAAGTGA 1180
 QY 1211 gtgagagtgaagaatgaagg 1234
 DB 1181 GCTGTGTGTTGTTCTAAGCAATGG 1204

RESULT 12
 X24873
 ID X24873 standard; cDNA; 1626 BP.
 AC X24873;
 DT 21-JUN-1999 (first entry)
 DE Medium-chain UDP-glucose:fatty acid glucosyltransferase cDNA.
 KW UDP-glucose:fatty acid glucosyltransferase; tomato;
 KW 1-O-acyl-a-glucose; acyl donor; esterification; ss.
 OS Lycopersicon pennellii.
 OS Lycopersicon esculentum.
 FH Key Location/Qualifiers
 FT CDS 1..1416
 FT /*tag= a
 FT W09909144-A1.
 PN 25-FEB-1999.
 PD 10-AUG-1998; IBI369.
 PF 30-JUN-1998; US-106464.
 PR 13-AUG-1997; US-055554.
 PA (CORR) CORNELL RES FOUND INC.

PI Banneatta NT, Ghangas GS, Kual J, Steffens JC;
DR WPI: 99-190155/15.
DR P-PSDB; W98009.
PT New chain length specific uridine diphosphate-glucose:fatty acid
PT glucosyltransferases - useful for preparing straight or branched,
PT (unsaturated 1'-O-acyl-a-glucoses and 1'-O-acyl-a-D-glucoses
PS Example 3; Page 33-36; 42pp; English.
CC The present sequence encodes a UDP-glucose:fatty acid
CC glucosyltransferase that transfers glucose to medium chain length
CC fatty acids to form 1'-O-acyl-a-glucoses. The 1'-O-acyl-a-glucoses
CC act as acyl donors in the esterification of glucose and in the
CC further esterification of partially acylated glucose in reactions
CC catalysed by glucose acyltransferases. They also act as acyl donors
CC in the esterification of sucrose and in the further esterification
CC of partially acylated sucrose in reactions catalysed by sucrose
CC acyltransferases. The source of the present cDNA is leaf trichome
CC mRNA of an F1 population of a cross between wild tomato
CC Lycopersicon pennellii (LA 716) and cultivated tomato
CC following immunoscreening of a trichome cDNA library. The
CC invention provides chain length-specific UDP-glucose:fatty acid
CC glucosyltransferases and a method for their purification. The
CC enzymes have specificity for transferring glucose to short, medium
CC or long chain length fatty acids. Methods for preparing
CC 1'-O-acyl-a-D-glucoses, where the acyl group comprises 3-5, 6-13
CC or 14-22 C atoms, are claimed.
SQ Sequence 1626 BP; 540 A; 256 C; 372 G; 458 T;

Query Match 4.7%; Score 70; DB 1; Length 1626;
Best Local Similarity 49.8%; Pred. No. 2.6e-09;
Matches 221; Conservative 0; Mismatches 205; Indels 18; Gaps 1;
QY 791 gaggagaataactcgttgagtggtgaactcgaagccgaatcttcggtggtatgtg 850
DB 793 GATGACACGCGTGTCTGAATGCTCTGATTCGAAGAAATCAAGTTCCATGTTATGTT 852
QY 851 tctgttgagcggttttgaggtttccaaagccacaaatggaagatttggaagggcta 910
DB 853 TGTGTTTGAAGTACAGCAGATTTTCATACAGCAGATGCAAGAACTTCTGATGGGCTA 912
QY 911 ttgacctcggaagcgcccttttatgatgatacagaagaatgacacgcgcgaa 970
DB 913 GAAGCCTCTGGGAAAGATTTTCATTTGGGTATACAGACAGGGAATGAAGATTGGTCCCA 972
QY 971 gaagaagaagaagaagagtgatgattggtggaattgaaataaaggggaaata 1030
DB 973 GAAGGATTGAGGAAAGAAC-----AAAGAAAAGGTTTATCATCA 1014
QY 1031 gtgtcgtggtcgtcagttggaggttctggcgccacctcgttgggagtgttcgtgacg 1090
DB 1015 AGAGGATGGGACCCCAAGTCTGATCTTGATCAGCAAGCTATTGAGCTTTTGTACT 1074
QY 1091 cattgtgggtgaaactcgtgtgtgagagcttgagttgcgggattccggtggtgcggtg 1150
DB 1075 CATTTGGGATGGAACTGCACACTGGAGGAATATACACAGGGGTACCAATGTTGACATGG 1134
QY 1151 ccgcagtggttgatcagacacgaatgcgaagctgattgagatcgtggtgggacaggg 1210
DB 1135 CCAGTATTTCGGGAACAGTTTTTCAATGAGAAGTTGCTGACTGAGGTAATGAGAAAGTGA 1194
QY 1211 gtgagagtgagaatgaatgaagg 1234
DB 1195 GCTGGTGTGTTCTTAACCAATGG 1218

RESULT 13
V23108
ID V23108 standard; cDNA to mRNA; 1597 BP.
AC V23108;
DT 06-AUG-1998 (first entry)
DE UDP-glucose:flavonoid 3,5-O-glucosyl transferase gene.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;

KW blue flower; ds.
OS unidentified.
FH key Location/Qualifiers
FT CDS 39..1460
FT /*tag= a
J10113184-A.
PD 06-MAY-1998.
PF 15-OCT-1996; 272253.
PR 15-OCT-1996; JP-272253.
PA (TOLG) TOKYO GAS CO LTD.
DR WPI: 98-315473/28.
DR P-PSDB; W56451.
PT Gentian flavonoid glucosyl transferase gene - used to produce
PT recombinant rose plants with blue petal(s)
PS Claim 3; Pages 7-8; 10pp; Japanese
CC The present sequence encodes a UDP-glucose:flavonoid 3,5-O-glucosyl
CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (sic) petals. The gene is
CC used to transform rose plants to produce blue flowers.
SQ Sequence 1597 BP; 483 A; 272 C; 368 G; 474 T;

Query Match 4.5%; Score 66.4; DB 1; Length 1597;
Best Local Similarity 50.0%; Pred. No. 2.3e-08;
Matches 237; Conservative 0; Mismatches 216; Indels 21; Gaps 2;
QY 778 ttctgaaaatacggagagagataactcgtggtgagtggtgaactcgaagccgaatcttc 837
DB 776 TATCGAAAGCCTGGAACCGAGAGATTTTGAATGGCTTGATATGAGCTGAGAAGTC 835
QY 838 ggtggtgtagtgcgtttggtgagcggttttgaggtttccaaagccacaaatggaagat 897
DB 836 TGTAGCTTTTCCCTTTGTTTGGTAGCCTTGGTCATTTTGGTAGGCCCAAGTAGGNAAT 895
QY 898 tgggaagggcttagctcgtcggaagggcccttttatgtagatgacagaaaca----- 951
DB 896 CGCATATGACATAGAGGGTAGTGGCCATAGATTTCTTGTGTCATTAAGAAGCCACCACC 955
QY 952 -----gaagaatgacgcgcgcgaagaagaagaagaagaagattgagttgagttg 1005
DB 956 TTTAGGGAATTTGAAGCCCTTGGTAGTATGAGAACTTGAAGAGGTCCTCCAGAGG 1015
QY 1006 ggaattgaaa-----aaaatggggaataagtcgtggtcgtcagttgaggt 1056
DB 1016 GTTTTATAGAACACACAGCTATACACCGAATGCTGATTTGGATGGCCCCACAGAGCGCGT 1075
QY 1057 tctggcgacctcgttgggagtttctgtagcagcttgggtggggaactcgcgttgga 1116
DB 1076 GTTCTCTCACTCAGCTGGGGAGGTTTCGTGTCGATTTGCGATGGAATTCACACTGGA 1135
QY 1117 gacgttgattcgggattcccggtggtggtgcggtgcgagttggttatcagacagcaga 1176
DB 1136 AAGCATTTTGGTTTGGTGTCCAGTGGCAACCTGGCCCTCTGTTGCTGAGCAGCATGAA 1195
QY 1177 tgcgaagcgtgattgagatgcgtgggacaggggtgagagtgagaatgaatga 1230
DB 1196 TGCATTTGAATTTGGTGAAGGAATTTGGCTTGGCTGAGAGGTTAAGATGGAATTA 1249

RESULT 14
T66178
ID T66178 standard; cDNA to mRNA; 940 BP.
AC T66178;
DT 15-JUL-1997 (first entry)
DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGU3-22.
KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
OS Brassica napus cv. Westar.
PN EP-771878-A1.
PD 07-MAY-1997.
PF 31-OCT-1995; 402425.
PR 31-OCT-1995; EP-402425.
PA (CAN) NAT RES COUNCIL CANADA.

PA (PLBZ) PLANT GENETIC SYSTEMS NV.
PI Grootwassink JMD, Hemmingsen SM, Kolenovsky AD, Peferoen M;
PI Reed DW, Underhill EW, Van Audenhove K;
DR WPI; 97-247418/23.
PT Plants genetically transformed to interfere with
PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression
PT - useful for production or rapeseed oil with reduced glucosinolate
PT content
PS Example 2; Page 20-21; 35pp; English.
CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase
CC (S-GT) partial cDNA clones pGL3-22 (T66178) and pGL4-2 (T66179)
CC were obt'd. by PCR-RACE (see also T66175-77) of B. napus cDNA.
CC The PCR product in pGL4 is completely contained in that of pGL3.
CC pGL3 includes 982 bp of s-gt open reading frame, of which the last
CC 116 bp overlap the g-qt fragment cloned in pGL2 (see also T66173-
CC 74). A full-length cDNA clone (T66166) for B. napus S-GT (W09825)
CC was subsequently obt'd.
SQ Sequence 940 BP; 209 A; 285 C; 209 G; 237 T;

Query Match 4.3%; Score 63.8; DB 1; Length 940;
Best Local Similarity 45.4%; Pred. No. 9.2e-08;
Matches 402; Conservative 0; Mismatches 462; Indels 21; Gaps 4;

QY 74 cacataaaccgcccctcccaattccgaagagactcctaaagccggaactgacg 133
DB 1 CACCTCAACCCATGTCGTCAGTTCGCTAAACGCTAGTCTCCAAAGGCTCAAGTCA 60

QY 134 ttttccagagcgtttatgatgagcgccgcatggcgaacacacagccctccgctgcccga 193
DB 61 ATCGCCACACACCTACACCGCTCTCTCCATCTCCACCCCTCCGCTCGAAGCA 120

QY 194 aacccacgggctcgaactcgtgggttcctccagcggtacagagagggctgaagccc 253
DB 121 ATCTCCGACGGCCACGACTCATCTCCCATAGGCGTCCCGGGCTCAGCATCGACGCTAC 180

QY 254 ggcggcagggagcgtacatgctccagatgaagccgcggctccgagggccttaaga 313
DB 181 TCCGAATCTTCAAGCTCCACGGCTCCCAACCTTAACCGCGTAACTCCCAATTCAAA 240

QY 314 aacctctctcaaacagcagcgtcaattgcttgctactcctccactctttgcatgg 373
DB 241 TCCACAGATCCCCCATCGA-----TCTTTAGTACGACTCTTCTCTCCCGTGG 291

QY 374 gcggcgagggcggtttgtccacgctccgacgcgcctctctggtgagccgcgc 433
DB 292 GGACTCGAAGTCGGGAGATCAACTCCCTCTCAGCTGCCGCTTTCTTCCCAACCACTC 351

QY 434 accgtgctgtg---catataccacttctactcaacggtacgcagagagatcgagcc 490
DB 352 ACCGTTTCTCTGCTCTCCGAATTCGGCTCCGGTGAGTTTCTCTCCCGCTGATCCC 411

QY 491 gtttccatgaattcagctcctccggttccatccctggagcagcgagttctccagc 550
DB 412 GCTTCCGGCTGTATCTCTGCTCGCTGGCTTCCCGCTTTAGCTACGAGGAGTCTCTTCC 471

QY 551 tttctgctcgtcgcagcgcggagagattccggttgatgatgaagagagagctggaact 610
DB 472 TTCGTGGGCTGCTACTGCTGACCCAGCGGACCGAGAGTCTTCTGAACCACTTC 531

QY 611 ttacagcgttgaagagagcgaagtattggtgaacacgtttgatcgcttggagccoga- 669
DB 532 ATTAACCATGAAGATGCTGATGCTGCTTCCGTCACAGGCTTCGAAGGGTTAGAGACAA 591

QY 670 -----tgcactcaacgctattgattgattgattgattgattgattgattgattccc 724
DB 592 GGTTCGTAAGTGGAGATCAGAGATATGAAGGCGACGCTTATCGGACCTATGATCCCA 651

QY 725 tccgctcttggagcggcgaagatccctccgaaagcttctacggcggcgatcttttoga 784
DB 652 TCTGCTTATCTGACGCCCAATCAAGACGATAAAGGCTACGGCTCAGCTGATG--- 708

QY 785 aaatcggaggagaataaactcgtggagtggttggtgaactcgaagcgaactcttcgtggtg 844

DB 709 AAGCGCTCTGGGAGGAGTGTATGGAGTGTATAGACACTAAGCTGAGTAAGTCGGTGT 768
QY 845 tatgtgctgtttgggagcgttttgaggtttccaaagccacaaatggaagagattgggaa 904
DB 769 TTTGTTTGGTTGGTTCTCTTTAGGATCCCTTTTGAGAACCAACTAGCTGAGTAGCAACG 828
QY 905 gggctattagcctcgggaagccctttttatggatgatacgagaa 949
DB 829 GCGTTACAGAAGAACCTTAACCTTCTTGTGGTGTGATTAAAGAA 873

RESULT 15
Q74684
ID Q74684 standard; cDNA; 1650 BP.
AC Q74684;
DT 22-JUN-1995 (first entry)
DE Early Ripening Tomato protein ERT1b gene.
KW Early Ripening Tomato; transgenic plant; crop improvement;
KW fruit ripening; transformation; ss.
OS Lycopersicon esculentum.
PN W09421794-A.
PD 29-SEP-1994.
PF 22-MAR-1994; G00581.
PR 22-MAR-1993; GB-005860.
PR 22-MAR-1993; GB-005862.
PR 22-MAR-1993; GB-005865.
PR 22-MAR-1993; GB-005859.
PR 22-MAR-1993; GB-005866.
PR 22-MAR-1993; GB-005867.
PR 22-MAR-1993; GB-005868.
PR 12-JUL-1993; GB-014351.
PR 12-OCT-1993; GB-020586.
PA (ZENE) ZENECA LTD.
PI Abu-Bakar UK, Barton SL, Gallego-Veigas PP, Gray JE;
PI Grierson D, Lowe AL, Picton S, Whotton LC, Abubakar UK;
DR WPI: 94-317014/39.
PT DNA constructs encoding fruit-ripening related proteins - useful
PT for transformation of plants to modify fruit quality
PS Claim 1; Page 47; 74pp; English.
CC ERT1b homologous transcripts are 1.8 kb in size and it is only
CC expressed during ripening of the wild-type fruit. Its expression is
CC highest during early stages of fruit ripening (e.g. breaker plus 3).
CC Levels of expression of the mRNA are low in rin (ripening inhibitor)
CC mutant tomatoes and are restricted to ripening rin tomatoes. The
CC gene is not activating upon wounding. The sequence has been
CC deposited as NCIMB 40544. The DNA may be used to transform plants
CC for production of plants with an altered level of expression.
CC Improvements in transgenic plants include resistance to damage and
CC pests, longer shelf life, improved flavor/aroma, etc.
SQ Sequence 1650 BP; 492 A; 338 C; 348 G; 472 T;

Query Match 4.0%; Score 58.6; DB 1; Length 1650;
Best Local Similarity 48.6%; Pred. No. 2.8e-06;
Matches 191; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 791 gaggagaataactcgtggagtggttgaaactcgaagccgaatacttcggtggtatgtg 850
DB 732 GAAGACAATAGCTGCATCGAGTGGCTCGATACACAAGCCCCCTAACTCTCTCTATGTC 791
QY 851 tcgtttgggagcgttttgaggtttccaaagccacaaatggaagagattgggaagggcta 910
DB 792 AGCTTGGGTAGCCTAGTAGGATTGATCAAAAGACTTGTATGACACTGCTTGGGATTA 851
QY 911 ttacgctcgggaagcccttttatggatgatacgagagacagaagaatgagacggcgaa 970
DB 852 GCTAATACGATCAACCGTCTCTGTGGTTATTTCG---ACCTGGCTCTGCTCTGGCTTT 908
QY 971 gaagaagaagaagaagagtgtagttggtgggaattgaaaaaataatggggaataa 1030
DB 909 CAATCTGCTGAGGCACTGCCTGATGTTTGTGAAAAATGTTAGAAAAATGTTAGGAGAAAGAGACGAATA 968

THIS PAGE BLANK (USPTO)

Result No.	Query Match	Score	%		DB	ID	Description
			Match	Length			
1	121.6	8.2	1731	4	US-08-466-593-1	Sequence 1, Appli	
2	121.6	8.2	1731	6	PCT-US95-07820-1	Sequence 1, Appli	
C	3	84	7218	1	US-08-232-463-14	Sequence 14, Appli	
	4	7	1627	5	US-09-106-464-1	Sequence 1, Appli	
	5	3.9	1669	3	US-08-522-421-1	Sequence 1, Appli	
6	51.2	3.5	1738	3	US-08-379-482A-2	Sequence 2, Appli	
7	49.8	3.4	1607	4	US-08-797-226-1	Sequence 1, Appli	
8	48.6	3.3	405	4	US-08-975-316-32	Sequence 32, Appli	
9	48	3.3	28958	1	US-08-258-261B-6	Sequence 6, Appli	
10	48	3.3	28958	1	US-08-458-837-6	Sequence 6, Appli	
11	48	3.3	28958	1	US-08-457-342-6	Sequence 6, Appli	
12	48	3.3	28958	1	US-08-457-646A-6	Sequence 6, Appli	
13	48	3.3	28958	2	US-08-458-076A-6	Sequence 6, Appli	
14	48	3.3	28958	2	US-08-764-233A-4	Sequence 4, Appli	
15	48	3.3	28958	2	US-08-457-335A-6	Sequence 6, Appli	
16	48	3.3	28958	2	US-08-723-214-6	Sequence 6, Appli	
17	48	3.3	49377	2	US-08-764-233A-1	Sequence 1, Appli	
18	47.6	3.2	2064	1	US-08-343-428-1	Sequence 1, Appli	
19	43.8	3.0	13987	3	US-08-804-227C-13	Sequence 13, Appli	
20	43.8	3.0	44377	3	US-08-804-237C-7	Sequence 7, Appli	
21	43.8	3.0	44377	3	US-08-804-198-1	Sequence 1, Appli	
22	42.2	2.9	2589	4	US-08-482-728A-3	Sequence 3, Appli	
23	41.4	2.8	1734	7	5352575-8	Patent No. 5352575	
24	41.4	2.8	2336	6	PCT-US92-00282-1	Sequence 1, Appli	
25	41.4	2.8	2339	6	PCT-US92-00282-2	Sequence 2, Appli	
26	40.8	2.8	30001	1	US-08-125-468-1	Sequence 1, Appli	

Db	1433	GTCTGTC 1440	11	1	1
<p>RESULT 3</p> <p>US-08-232-463-14/c</p> <p>Sequence 14, Application US/08232463</p> <p>Patent No. 5670367</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: DORNER, F.</p> <p>APPLICANT: SCHEIFLINGER, F.</p> <p>APPLICANT: FALKNER, F. G.</p> <p>TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS</p> <p>NUMBER OF SEQUENCES: 52</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Foley & Lardner</p> <p>STREET: 1800 Diagonal Road, Suite 500</p> <p>CITY: Alexandria</p> <p>STATE: VA</p> <p>COUNTRY: USA</p> <p>ZIP: 22313-0299</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patentin Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION DATA: US/08/232,463</p> <p>FILING DATE:</p> <p>CLASSIFICATION: 435</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/07/935,313</p> <p>FILING DATE:</p> <p>APPLICATION NUMBER: EP 91 114 300.6</p> <p>FILING DATE: 26-AUG-1991</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: BENT, Stephen A.</p> <p>REGISTRATION NUMBER: 29,768</p> <p>REFERENCE/DOCKET NUMBER: 30472/114 IMMU</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (703)836-9300</p> <p>TELEFAX: (703)683-4109</p> <p>TELEX: 899149</p> <p>INFORMATION FOR SEQ ID NO: 14:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 7218 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>IMMEDIATE SOURCE:</p> <p>CLONE: pTZgpt-F1s</p> <p>US-08-232-463-14</p>					
<p>Query Match 5.7%; Score 84; DB 1; Length 7218;</p> <p>Best Local Similarity 2.3%; Pred. No. 1.2e-13;</p> <p>Matches 9; Conservative 250; Mismatches 125; Indels 0; Gaps 0;</p>					
QY	934	atggatgacagacagacgaagaatgacagcgagcgaagaagaagaagaagaagtt 993	11	1	1
Db	1451	ATAGAGAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392	11	1	1
QY	994	gagttgattgggggaattgaaaaaatggggaaaatgtctgtgtcgcagtga 1053	11	1	1
Db	1391	RR 1332	11	1	1
QY	1054	ggttctggcgaccctcggtggatgtttcgtgacacattgggtggaactcgctgt 1113	11	1	1
Db	1331	RR 1272	11	1	1
QY	1114	gagagcttgagttcgccggattccggtggtggtgcgcagttgattgacagac 1173	11	1	1
Db	1271	RR 1212	11	1	1

Db 853 TGGTTTGGAGTACAGCAGATTTCACTACGACACAGATGCAAGAACTTGCATATGGGCTA 912
Qy 911 ttacccctcggaagccctttttatgatgacgagacagaaatgacgagcgaa 970
Db 913 GAAGCCCTCGACAAAGATTTCAATTTGGTGTATCAGAACAGGGAATGAAGATTTGCTCCCA 972
Qy 971 gaagaagaagaagaagaagattgattcattgggggaattgaaaaaaatggggaata 1030
Db 973 GAAGGATTCGAGGAAGAAC-----AAAAAGAAAAGGTTTAAATCAT 1014
Qy 1031 gtgcgtggtgctgcagtgagggttctggcgacacccctggttgggagtttccgtgacg 1090
Db 1015 AGAGATGGGCACCCCAAGTGCTGATTCTTGATCAGGAAGCTATTGGAGCTTTTGTACT 1074
Qy 1091 catgtggtgggaactcgtgtggagagcttgagttcgggattcccggtggtgagcg 1150
Db 1075 CATGTGTGATGGAACCTCGACACTGGAAGGAATATCAGCAGGGGTACCAATGTTGACATGG 1134
Qy 1151 ccgagtggtttgatcagacgcaatgcgaagctgattgaggaagcgtgggggacaggg 1210
Db 1135 CCAGTATTTCGGGAACAGTTTTCATGAGAAAGTTGGTGACTGAGGTAATCAGAAAGTGA 1194
Qy 1211 gtgagagtggaatgaatgaagg 1234
Db 1195 GCTGTGTGTGTTCTTAAGCAATGG 1218

RESULT 5

US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
; GENERAL INFORMATION:
; APPLICANT: Abu-Bakar, Umi Kalsom
; APPLICANT: Barton, Sarah Louise
; APPLICANT: Gallego-Veigas, Pedro Pablo
; APPLICANT: Gray, Julie Elizabeth
; APPLICANT: Grierson, Donald
; APPLICANT: Lowe, Alexandra Louise
; APPLICANT: Picton, Steve
; APPLICANT: Whorton, Lee Collin
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS
; TITLE OF INVENTION: DERIVED THEREFROM
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/522,421
; FILING DATE: 11-JAN-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305868.3
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305869.1
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305859.2
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305865.9
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9305866.7
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305867.5
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305860.0
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305862.6
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9314351.9
; FILING DATE: 12-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9320988.0
; FILING DATE: 12-OCT-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: ERTIB
US-08-522-421-1

Query Match 3.9%; Score 57; DB 3; Length 1669;
Best Local Similarity 48.3%; Pred. No. 1.7e-06;
Matches 190; Conservative 0; Mismatches 200; Indels 3; Gaps 1;
Qy 791 gagagagaataactcgtgagtggttgaactcgaagccgaaatcttcggtggtgatgtg 850
Db 732 GAAGCAATATAGCTGCATCGATGCTCGATAGACAAAGCCCTTAACCTCTGTTCTATGTC 791
Qy 851 tctgttggtgagcgttttgaggtttccaaagggcacaataaggaagattgggaagggccta 910
Db 792 AGCTTGGGTAGCCTAGTGAGGATTGATCACAAGAGAGTTGATTGAGACTGCTTGGGATTA 851
Qy 911 ttacgtcgtgggaagggccctttttatggatgatacagagaacagaagaatgacgagcgaa 970
Db 852 GCTAATAGCGATCAACCCGTTCTTGGGTTATTTCG---ACCTGGCTCTGCTCTGGCTTT 908
Qy 971 gaagaagaagaagaagaagattgagttggtgggaattgaaaaaatggggaata 1030
Db 909 CAATGCTGCTGAGGCACCTGCTGATGTTTGGAGAAAATGGTAGAGAAAAGAGGACGAATA 968
Qy 1031 gtgcgtggtgctgcagtggtgaggttctggcgacacccctgcttgggatgtttcgtgacg 1090
Db 969 GTGAATGGGCACCACAAACAGGTGCTTGACATCCCGCGGTAGCAGGGTTTTCCTACT 1028
Qy 1091 cattgtggtggaactcgtgtggagagcgttgagttgcggattcccggtggtgagcggtg 1150
Db 1029 CATTTGTTGTAATTTACGCTTGAAGTATATGTGAAGAAGTCCCTATGTTGTCGACAG 1088
Qy 1151 ccgagtggtgttgatcagacgagcaatgcaag 1183
Db 1089 CCATTTCTAGCAGACCAACTGGTGAACGCAAG 1121
RESULT 6
US-08-379-482A-2
; Sequence 2, Application US/08379482A
; Patent No. 5859334
; GENERAL INFORMATION:
; APPLICANT: Brugliera, Filippa
; APPLICANT: Holton, Timothy A.
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING
; TITLE OF INVENTION: GLYCOSYLTRANSFERASE ENZYMES AND USES THEREFOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

QY 807 tggagtggtgaactcgaagccgaacacattcttcggtggtatgtatgtcgttttgggagcgctt 866
Db 852 TAGATTGGTTGAATGCACACAAACCTTAAATCGGTTCTCTATGTATCTTTTCGGAAGCATGG 911
QY 867 tgggtttccaaaggccacaaatggaagagattgggaagggtctattagcctgcggaagcc 926
Db 912 CTAGATTTCCTGAGAGCCCACTGAATGAATAGCCCAAGCTCTGGATCTTCA--AATGT 969
QY 927 cttttttatggtatgatacagagacgaagaatgacagcgcgagagagaagaagaag 986
Db 970 TCCCTTTCATTTTGTATGTAGGCGCTTAATGAAGAAACGGCGTTCGTGGTTGCCAGTTGTAA 1029
QY 987 aagagtgagtgctcattgggggaattgaaaaaaatggggaataatgtagtgcgtgggtgcctgc 1046
Db 1030 TTTAGAGGACAAGACT-----AAAAGGGTTTGTACATCAAAAGGTTGGTCCAC 1079
QY 1047 agtgagggttcggcgaccctcgcttgggtatgttttgcgcagcattgtgggtgggaact 1106
Db 1080 AGCTTACATCATGGAACATTCAGCAACAGCGGGTTTCATGACTATTTGGTACTAATT 1139
QY 1107 cggctgaggagagcttgagttgaggattccggtggtggtgcgcagtcggtgttgc 1166
Db 1140 CGGTTCTGGAAGCCATCCTTTTGGCGTCCCAATGATAACATGGCCACTTTTATGCTGATC 1199
QY 1167 agacagcgaatgcaagctgattgaagatgcgtgggg-----acaggggtgagagtga 1220
Db 1200 AATTCTACAGAGAGGTAGTCGAGGTTAGGGGATTGGGAATCAAAATCGGATAGATG 1259
QY 1221 gaatgaatgaaggggtgggttggatggatggtgagatgagaag 1263
Db 1260 TATGGAATGAAGGATTGAGATCAGCGGCGCTGTAATAGMAAG 1302

RESULT 8

US-08-975-316-32
; Sequence 32, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SLEATH, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-975-316-32

Query Match 3.3%; Score 48.6; DB 4; Length 405;

Best Local Similarity 49.2%; Pred. No. 0.00019;

Matches 157; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 789 cggaggagaataactcgcgtggagtggttgaactcgaagccgaatacttcggtggtgctatg 848
Db 88 CAGAAGACCCGAGTTGCTATTTCTCGCTCGACAAAGCCCTAACTCAGTGTCTATG 147
QY 849 tctcgtttgagagcgttttgaggtttccaaaggcacaatggaagagattgggaagggc 908
Db 148 TGAGTCTGTGGAGATCGCCTCTGTGAACGAGTCGGAATTTTCCGAATAGCTTTAGGTT 207
QY 909 tattagcctcggaagggccctttttatggtatgatacagagaacagaagaatgacgagcg 968
Db 208 TAGCCGATAGCCAGCAGCCATTCTTGTGGTGTTCG---ACCCGGGTCACTGAGCGCT 264
QY 969 aagagaagaagaagaagaagagtgagttgcattggggaattgaaaaaatgggaaa 1028
Db 265 CGGAACCTCTTAGAGAATTTGCCCGGTTGCTTCTGAGGCATTACAGGAGAGGGGAAGA 324
QY 1029 tagtgcgtggtgcgtgcagtgagggttctggtgcacccctgcgttgggattgttcgta 1088
Db 325 TTGTCAAAATGGCGCCTCAACATGAAGTCTGCTGCCTCATCGGCTCTCGGAGCGTTTGA 384
QY 1089 cgcattgtggtggaactc 1107
Db 385 CTCACAATGGATGAACCTC 403

RESULT 9

US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott

```
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-258-261B-6

Query Match          3.38; Score 48; DB 1; Length 28958;
Best Local Similarity 46.4%; Pred. No. 0.0022;
Matches 156; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 156 ggccgcatggcccaacacagctcccgccgctgcggaacccaccgagcctcgacttgcg 215
DB 10208 GCCACAGCCAGCGGAGATCCCGCGCTTCGTCGAGGCGTCTCTCCTCGAGGAGC 10267

QY 216 tggcgttcgcagcggtacagacagcggcgtgaagcccgccgagcgggagcgctaca 275
DB 10268 CGGCGCGCATCGCCCGCTGCGGAGAAAGCGCTACACCGTCGCGGCGGAGCGCA 10327

QY 276 tgcgcgagatgaagcccgccgctgcgagggccttaagaacccctctcaaacagcgc 335
DB 10328 TGGCGCGGTGCGGTGCGCGCTCCGACCTCCAGACCTACCTCTCCTCGGCGGACA 10387

QY 336 acgtcaacttgcgtctactccacccctcttgatggcgagcgcgcgcgcgcgcgcgcgc 395
DB 10388 GCGTCTCCACCGCGCGCTGCTGACGCTCTCACCAGGAGGCTACCTCTGATCCGCGGAGCGCG 10447

QY 396 cccacgtcccgaccgcccctcttggtcgagcccgccacccgctgtgtcatataccact 455
DB 10448 CGGTGACGCGGTGCTGACGCTCTCACCAGGAGGTTGTCGCCCGCAAGATCCGCG 10507

QY 456 tctacttaacggctacgcagacagatcgacgcgcg 491
DB 10508 TCGACTACGCTCCCACTCCCGCCAGATGAGACGCCG 10543
```

```
RESULT 10
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ring
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

Query Match          3.38; Score 48; DB 1; Length 28958;
Best Local Similarity 46.4%; Pred. No. 0.0022;
Matches 156; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 156 ggccgcatggcccaacacagctcccgccgctgcggaacccaccgagcctcgacttgcg 215
DB 10208 GCCACAGCCAGCGGAGATCCCGCGCTTCGTCGAGGCGTCTCTCCTCGAGGAGC 10267

QY 216 tggcgttcgcagcggtacagacagcggcgtgaagcccgccgagcgggagcgctaca 275
DB 10268 CGGCGCGCATCGCCCGCTGCGGAGAAAGCGCTACACCGTCGCGGCGGAGCGCA 10327

QY 276 tgcgcgagatgaagcccgccgctgcgagggccttaagaacccctctcaaacagcgc 335
DB 10328 TGGCGCGGTGCGGTGCGCGCTCCGACCTCCAGACCTACCTCTCCTCGGCGGACA 10387

QY 336 acgtcaacttgcgtctactccacccctcttgatggcgagcgcgcgcgcgcgcgcgcgc 395
DB 10388 GCGTCTCCACCGCGCGCTGCTGACGCTCTCACCAGGAGGTTGTCGCCCGCAAGATCCGCG 10447

QY 396 cccacgtcccgaccgcccctcttggtcgagcccgccacccgctgtgtcatataccact 455
DB 10448 CGGTGACGCGGTGCTGACGCTCTCACCAGGAGGTTGTCGCCCGCAAGATCCGCG 10507

QY 456 tctacttaacggctacgcagacagatcgacgcgcg 491
DB 10508 TCGACTACGCTCCCACTCCCGCCAGATGAGACGCCG 10543
```

```
RESULT 11
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ring
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
```


Db 10388 GGCCTCCACCGCCGCGCTCAACAGCCCCAGGCTACCCCTGATCCGCGAGCCCGCG 10447
QY 396 ccaacgtccgacgccccttctggtcgagcccgccaccgctgctgtcatataccact 455
Db 10448 CCGTCGACGCGCTGCTCGACGCTCTCACCGCCACCCAGGNGTTTCGCCCGCAAGATCCGCG 10507
QY 456 tctacttaacggctacgcagacagatcgacgccc 491
Db 10508 TCGACTACGCTCCACACTCCGCCAGATGGACGCG 10543

RESULT 13
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458.076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-458-076A-6

Query Match 3.3%; Score 48; DB 2; Length 28958;
Best Local Similarity 46.4%; Pred. No. 0.0022;
Matches 156; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 156 ggcgcgcgatggccaacacagcctccgcgctgcggaacccccacccgggctcgactcg 215
Db 10208 GCCACAGCCAGCGGAGATCGCGCGCTTCGTGCGAGGCGCTCTTCCTCGAGGAGC 10267
QY 216 tggcgtctccgacgctacgacgagggctgaagcccgcgagcggaagcgctaca 275
Db 10268 CGCGCGGATGCCGCCCTGCGCAGGAAGCGCTCACACCGTCGGCGGCAACGGCGGCA 10327
QY 276 tgcgcgatgaaagcccgcgctccgagggccttaagaaacctcttctcaacaaacg 335
Db 10328 TGGCGCGCTCGAGCTCGCGCGCTCCGACCTCCAGACTACCTGCTCCCTGGGCGACA 10387
QY 336 agtcactttcgtctactccacacctcttgcattggcgagggcgaggtgagcgcttgt 395
Db 10388 GGCTCTCCACCGCCGCGTCAACAGCCCCAGGGTACCTCGTATCCGCGGACCCCGCG 10447
QY 396 ccaacgtccgacgccccttctggtcgagcccgccaccgctgctgtgcataccact 455
Db 10448 CCGTCGACGCGCTGCTCGACGCTCTCACCGCCACCAAGGTGTTTCGCCCGCAAGATCCGCG 10507
QY 456 tctacttaacggctacgcagacagatcgacgccc 491
Db 10508 TCGACTACGCTCCACACTCCGCCAGATGGACGCG 10543
RESULT 14
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Szeanna
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764.233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Search completed: August 1, 2000, 18:16:29
Job time: 74238 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 18:20:52 ; Search time 51.03 seconds
(without alignments)
557,848 Million cell updates/sec

Title: US-09-147-955-2
Perfect score: 2421
Sequence: 1 MVRRLVLLATPPAQGHINPA.....GRDGSSLNLAFLHQVARA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_64: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match Length	ID	Description	
1	1053.5	43.5	467 2 T02238	glucosyl transferase	
2	1038	42.9	456 2 C71402	probable glucosylit	
3	888	36.7	458 2 C71420	hypothetical prote	
4	762.5	31.5	471 2 A54739	indole-3-acetate b	
5	702	29.0	479 2 E71419	probable indole-3-	
6	675.5	27.9	460 2 T00639	hypothetical prote	
7	663.5	27.4	456 2 T00506	indole-3-acetate b	
8	658.5	27.2	484 2 D71419	probable indole-3-	
9	653.5	27.0	438 2 T00507	indole-3-acetate b	
10	628.5	26.0	475 2 F71419	probable indole-3-	
11	604.5	25.0	453 2 T00511	indole-3-acetate b	
12	501.5	20.7	453 2 T45603	glucosyltransferas	
13	484	20.0	455 2 T00584	indole-3-acetate b	
14	466.5	19.3	451 2 T45604	glucosyltransferas	
15	458.5	18.9	452 2 T12981	hypothetical prote	
16	455.5	18.8	440 2 T00583	indole-3-acetate b	
17	455	18.8	438 2 T45602	glucosyltransferas	
18	445	18.4	449 2 T45605	glucosyltransferas	
19	444.5	18.4	447 2 T12978	hypothetical prote	
20	414.5	17.1	462 2 T0732	UTP-glucose glucos	
21	413	17.1	347 2 T06371	probable UDP-glucu	
22	407.5	16.8	452 2 T00981	flavonol 3-O-glucu	
23	406.5	16.8	466 2 T07404	probable glucosylit	
24	406.5	16.8	471 2 S01037	flavonol 3-O-glucu	
25	403.5	16.7	385 2 T12980	hypothetical prote	
26	400.5	16.5	476 2 T03745	glucosyltransferas	
27	398.5	16.5	420 2 T08005	flavonol 3-O-glucu	
28	398	16.4	433 2 S31767	glucosyl transfera	
29	398	16.4	452 2 G71416	probable glucosylit	

flavonol 3-O-glucu
flavonol 3-O-glucu
probable glucosylit
glucuronosyl trans
glucosyltransferas
UDP-glucose glucos
flavonol 3-O-glucu
UTP-glucose glucos
hypothetical prote
UTP-glucose glucos
UTP-glucose glucos
UTP-glucose glucos
UTP-glucose glucos
UTP-glucose glucos
glucosyltransferas

30 397.5 16.4 471 1 S08325
31 397.5 16.4 471 2 S01052
32 396.5 16.4 478 2 T05423
33 395 16.3 472 2 S39507
34 392.5 16.2 476 2 T03747
35 391 16.2 488 2 T07786
36 389.5 16.1 455 1 XUBHFG
37 384 15.9 487 2 S41951
38 379.5 15.7 478 2 A71417
39 367.5 15.2 394 2 S41952
40 367.5 15.2 478 2 T08395
41 363 15.0 490 2 T46162
42 361.5 14.9 346 2 S41954
43 358.5 14.8 481 2 T01850
44 357 14.7 287 2 S41953
45 356.5 14.7 507 2 T46161

RESULT 1
T02238
glucosyl transferase, jasmonate-induced - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 24-Nov-1999
C:Accession: T02238
R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.
submitted to the EMBL Data Library, January 1997
A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspensio
A:Reference number: Z14633
A:Accession: T02238
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-467 <KOJ>
A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901
A:Experimental source: strain BY-2
C:Genetics:
A:Gene: J1GT
C:Superfamily: flavonol O3-glucosyltransferase

ALIGNMENTS

Query Match		43.5%	Score 1053.5;	DB 2;	Length 467;
Best Local Similarity		45.3%	Pred. No. 3.4e-77;		
Matches 209;		Conservative 83;	Mismatches 150;	Indels 19;	Gaps 7;
Qy	6	VLLATPPAQGHINPALQFAKRLKAGTDVFTTSVVAWRMANTASAAAGNPPGLDFVAF	65		
Db	10	VLLATPPAQGHINPSPQFQSKKLINLGKVTLSLSLAPNKNLPKI-----EGLTAPF	64		
Qy	66	SDGVDDGLKPCGDKR-YMSEMKARGEALRNLL----LNHDVTFFVYSHLFAWAEEVA	120		
Db	65	SDGYDGNFKGDFDYYHLFNSAISKHSEPIANLIKSKARKNGYPFTRVITYTILMDAGSVA	124		
Qy	121	RESQVPSALLWSPATVLCIYFYFNGYAD---EIDAGSDEIQLPRLPQLPQSLPFL	177		
Db	125	KLHIPSTLFWIQPATVFDIYYRFTNFANYFNKNDSDQDIIEQLPSLSDFSPFVF	184		
Qy	178	PETERPRLM--MKELETLDGEEKAKVLVNTFDALPDALTAIDRYELIGLIPISAF	235		
Db	185	DDVKSNDWAVESIKROTEILNSENPRILVNTFDALNLALRVLNKNTVMVGIGLIPSSF	244		
Qy	236	LDGGDPSETSYGGDLPEKSENNCEWMLDTKPSVYVYSGSVLRFPPKAEIGKGL	295		
Db	245	LDEKDRKDNFADAMLE--SENNYMWLDRANKSVIYAFGSYAEISQWBEISGLL	302		
Qy	296	ACGRPFLWMTREQNDGDEEEELSCIGELKMKGIYVSWCSQLEVLALHPALGCVFTHCG	355		
Db	303	KCGRPFLWVIRETLN--GEKPEKLTCKDELEKIGRIVRWCSQMEVLKHSVGCFLTHCG	360		
Qy	356	WNSAVESLSGVPVAVPQWFDOTTNAKLIEDAWGTGVYVRMNEGGVSDGSEIERCEVEMV	415		
Db	361	WNSTLESASGVPIVACPIWNDQICNAKLIDQVWIKGVNRVANKEGIKRDFORCEIV	420		

A:Accession: A54739
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-471 <SZ>
A:Cross-references: GB:L34847; NID:g548194; PIDN:AAA59054.1; PID:g548195
C:Superfamily: flavonol O3-glucosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 31.5%; Score 762.5; DB 2; Length 471;
Best Local Similarity 39.1%; Pred. No. 9.9e-54;
Matches 187; Conservative 70; Mismatches 182; Indels 39; Gaps 16;

Qy 6 VLLATPFAOCHINPALQFAKRLKAGTDTVFTTSYAWRRMANTASAAAGNPPGLDFVAF 65
Db 5 VLVVPPGOGHNMVPMQFAKRLKAGTATLVTT-----RFQRTADVDA-HPAMVE-AI 57
Qy 66 SDGYDGLKPCGKG-KRYSEMKAARGSEALRNLL--LNNHDV-TFVYVSHLFAWAEVA 120
Db 58 SDGHDGEGFASAGVAEYLEKQAAAASLASLVEARASSADAFTCVYVDSYEDWVLPVA 117
Qy 121 RESQVPSALIMVPEPATVLCIYYFYFNG-----YADEIDAGSDEIQLPDPR---LPPLE 168
Db 118 RMGLPAVPFSTQSCAVSAVYHFSGRLAVPPGAADGSGGAGAAALSEAFGLPEME 177
Qy 169 QRSPLTFLPETPERFLMKKEKLETLGEEKAKVLVNTFDALPDALTADRY-ELIGI 227
Db 178 RSELPFVFDHGPYPTIANQAIKQFAHAKDD-WYLFNSFEETETEVLAGLTKARAI 236
Qy 228 GPLI--PSAPFDGDPSETSYGGDLFEKSENNCWEVLDTPKSSVYVYVSGSVLRFPKA 285
Db 237 GPCVPLTAGTAGANGRTYCANLVK--PEDACTKWLDTKPDRAVYVYVSGSVLSIGNA 294
Qy 286 QMEIGKGLLAGCPPLMIRQKNDGDEEBEELSICGLKMG--KIVSWCSQLEVL 343
Db 295 QKELARGLLAGKPLWVYRA-----SDEHQVPRYLAEATATGAAMVVPWCPOLDVLA 349
Qy 344 HPALGCFVTHCGMNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRMNEGGV 403
Db 350 HPVAGCFVTHCGMNSTLEALSGVPMVAMALWTDQPTNARVWELAGVARRDAGV 409
Qy 404 D-GSEIERCVEMVMDGSEKSLVRENATKWKTLAREAMGEDGSSLKMLNAPLHQVARA 460
Db 410 FLRGEVERCVRAVMDGGEAASAAKAAAGWRDRAAVAPGSSDRNLDEFV-QFVRA 466

RESULT 5
E71419
Probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: E71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis tha
A:Reference number: A71400; MUID:98121113
A:Accession: E71419
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-479 <BE>
A:Cross-references: GB:Z97339; NID:g2244901; PID:e326931; PID:g2244906
C:Genetics:
A:Map position:
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 29.0%; Score 702; DB 2; Length 479;

Best Local Similarity 34.9%; Pred. No. 7.7e-49;
Matches 166; Conservative 87; Mismatches 182; Indels 40; Gaps 14;

Qy 6 VLLATPFAOCHINPALQFAKRLKAGTDTVFTTSYAW--RRMANTASAAAGNPPGLDF 62
Db 9 VMLVSPFGOGHNPPLRLGLKLTASKGLLVTFVTTTEKPMGKMRQANKIQDGLVKPVGLGF 68
Qy 63 VA---PSDYDGLKPCGDKRYMSEMKARGSEALRNLL--LNNHDVTFVYVSHLFAWAA 117
Db 69 IREFEFDGFADDDEKRFDAFRPHLEAVGQKEIKNLVKRYNKEPVTCLINNAFVFWC 128
Qy 118 EVARESQVPSALIMVPEPATVLCIYYFYFNGVA-----DEIDAGSDEIQLPPLPLEORS 172
Db 129 DVABELHIPSVAVLWQSCACLTATYVYVHRLVKFPTKPEDI---SVEICPLPLKDEI 185
Qy 173 PTFLLPETP-ERELMKKEKLETLGEEKAKVLVNTFDALPDALTADRY---ELIGI 227
Db 186 PSLFHSPPPTAFGDIIQLQKRFENHKSFYLDITFRELEKIDMHHSQCLCPQAIISPV 245
Qy 228 GLTIPSAFLDGGDPSETSYGGDLFEKSENNCWEVLDTPKSSVYVYVSGSVLRFPKAQM 287
Db 246 GPLFKNAQTLSSDVK-----GDISEPA--SDCMEWLDSPSSVYVYISFGTIANLKQEQM 298
Qy 288 EETGKGLLAGCPPLMIRQKNDGDEEBEELSICGLKMGKTVSWCSQLEVLAPAL 347
Db 299 EETAHGVLSLGLSVLWVVRPPM--EGTFVPEHV-LPREEEKGLIVEMCPQERVLAHPAI 355
Qy 348 GCFVTHCGMNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGVRVRMNEGGVDSGE 407
Db 356 ACFLSHCGMNSTMEALTAGVPCPPQWGDQVTDVAVLADVFKVTGVL---GRGAAEEM 411
Qy 408 I---ERCVENMVDG--GEKSLVRENATKWKTLAREAMGEDGSSLKMLNAPLHQV 457
Db 412 IVSRVVAEKLEATVGERKAVELRENNARKAAEAADVAGGSSDMNFEKFDKL 466

RESULT 6
T00639
Hypothetical protein F316.2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00639
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Arau
; Vysotskaia, V.S.; Yu, G.; Becker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z14197
A:Accession: T00639
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-460 <PED>
A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829862; GSPDB:GN000059; ATSP:F3
C:Genetics:
A:Gene: ATSP:F316.2
A:Map position: 1
A:Introns: 219/3
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 27.9%; Score 675.5; DB 2; Length 460;
Best Local Similarity 34.9%; Pred. No. 1e-46;
Matches 169; Conservative 72; Mismatches 186; Indels 57; Gaps 15;

Qy 2 VRRVLLATPFAOCHINPALQFAKRLKAGTDTVFTTSYAWRRMANTASAAAGNPPGLD 61
Db 8 VKGHVILPVPVQGHLPWQFAKRLVSKNVKTATTYY-----TASSITTSLS 58
Qy 62 FVAFSDGYDGLKPCG----DGKRYMSEMKARGSEALRNLL-----LNNHDVTFVYVSHLF 113
Db 59 VEPISDGF--FIPIGIPGFSVDYTESSEFKLNGSETLLTLEKFKSTDPIDCLIIYDSFL 116
Qy 114 AWAAREAVAREQVPSALLWVEPATVLCIYYFYFNG-YADEIDAGSDEIQLPPLPLEORS 172
Db 117 PMGLEVARSMELSAASFFTNLTVCVSLKNSGDFPLPADPNSAPFRIGLPSLSDEL 176

Qy	366	GVPVAVPWFQDTTNAKLIBDANGTVGVVRMNEGQVGD	---	SEIERCEVMVMDGGEK	421
				: : : :	
Db	344	GVPVAYPTWIDOPDARLLVDVFCIGVRMK	--	NDAIDGELKVAEVCIEAVTEGPAA	400
Qy	422	SKLIVRENAIKWTKLAREAMGEDGSSGLKNLNAFLHGV			457
		:			: : : :
Db	401	ADM-RRRATELKHAARSAMSPGGSSAQNLDSTFSDI			435
RESULT	10				
F71419					
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana					
C;Species: Arabidopsis thaliana (mouse-ear cress)					
A;Variety: Columbia					
C;Date: 03-Aug-1998					
C;Accession: F71419					
#sequence_revision 03-Aug-1998 #text_change 24-Nov-1998					
R.;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks, P.; Wedler, H.; Wedler, E.; Wambolt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Gladstavanabh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Kieger, M.; Schaeffer, M.; Funk, B.					
Nature 391, 485-488, 1998					
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech					

```

118 EVARESQVPSALLWVEPAIVLCIYFYFNGYAIDEIDA-----GSDEIQLPRLPPLEQRS 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 AAAKECKLNNIIFTSTATACFASRVFDKLYANNVOAPLKETKGOOEELVPPEFYLRKYD 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 LPTFLLPETPERF-----RLUMKEKETLDEGEKAKVLNFTDALEPDLATDAIDRYEL---- 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 FPV-----SFAESLEIMEVYNTVDKRTASSVINTASCLESSLSFLOQQOQLIPV 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
225 IGTGPL-----IPSAFLDGGDPSETSYGGDLFEKSENNCNVEMLDTKPKSSVYVVGSV 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 YPIGPHVVASAPTSLL-----ENKSCIEWLANKQKVNYSYIISMGS 276
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280 LRPPKAQMEIEIGKLLACGRPFLMWIREQNKNDGDEEEELSCIGELKKM-----CKIVSW 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
277 ALMEINEIMEVASGLAASNQHLWVIRPGSTPGSEWIE---SMPEFSKMWLDRGYIVKM 333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 CSOLEVLAHPALGCFVTHCGWNSAVESLSGCVVPVAPQWFDDQTTNAKLIEDANQGTGVRV 395
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 APOKEYLSHPAYGGFWMSHCWNSTLESIGOGVPMICRPFSGDGQKVNARYLECVNKGIGTV 393
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
396 RMNEGGGVGDSIERECVE--VMVDDGGKSKLVRENAIKWTKLAREAMGDGSSLKNLNAF 453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
394 E-----GELDRGVVERAVKRLVMDEEGE---MRKRAFSLKEQLRASVKSGGSSHNSLEEP 446
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
454 LH 455
      : :
447 VH 448
      : :

RESULT 13
T00584
indole-3-acetate beta-glucosyltransferase homolog T27E13.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00584
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.

```

A:Description: Arabidopsis thaliana chromosome 11 BAC T2/E13 genomic sequence.
A:Reference number: Z14177
A:Accession: T00584
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-455 <R0U>
A:Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150406
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A: Introns: 166/1
A:Note: T27E13.12
C:Superfamily: flavonol O3-glucosyltransferase

Query Match	20.0%	Score 484;	DB 2;	Length 455;
Best Local Similarity	30.6%;	Pred. No. 2.7e-31;		
Matches 147;	Conservative 71;	Mismatches 196;	Indels 66;	Gaps 17;

QY	62	FVAPS---	DGYDDGURPGDGRKRTSEMKANGSALKNLL---	UNNHIDVITFVYVISHUFWA	116
Db	64	RIHFSTLPLNIPSELVRAKDFIGTIDAVYTRLEEPPEKLLDSUNSPSPSVIFADYVYVWA	123		
QY	117	AEVARESQVPSALLNVEPATVLCIY---	VYFNGYADEIDAGSDEI---	QLPRLPPLRQR	170
Db	124	VRGRKRNIPVYSLTMSNTILSFLHSDLLISHGHA--	LEPSEEEVVDVYVGLSPTKLR	182	
QY	171	SLPTFLPPTPRFRFLMKKEKLETLDGEERAKVUWNTFDALEP---	DALTATDRYELIGI	227	
Db	183	DLPP--TFDGYSDRVFTAKLCFDELPGAR--	SLFTTAYELEHKADAETSKLDIPVYAI	239	
QY	228	GPLIPSAFLDGDGPSETSYGDDGLFEKSEENNCVEMLDTRPKSSVYVYSGVYVREFPKQAM	287		

Search completed: August 1, 2000, 21:27:44
Job time: 11212 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 13:52:00 ; Search time 47.63 Seconds
(without alignments)
228.755 Million cell updates/sec

Title: US-09-147-955-2
Perfect score: 2421
Sequence: 1 MVRRLVLLATFPAQGHINPA.....GEGSSLKLNFLHGVARA 460

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2421	100.0	460	1 W92947	WO9905287 Seq ID 7
2	2293	91.8	443	1 W92948	WO9905287 Seq ID 8
3	1646.5	68.0	471	1 W92949	WO9905287 Seq ID 9
4	1445	59.7	468	1 W92950	WO9905287 Seq ID 1
5	1363.5	56.3	468	1 W92952	Maize UDP-glucose
6	762.5	31.5	471	1 W03756	UDP-glucose:thiohy
7	700	28.9	466	1 W09825	Medium-chain UDP-q
8	407.5	16.8	471	1 W98009	Glucosyl transfera
9	406.5	16.8	470	1 W47172	Solanum melongena
10	398	16.4	433	1 W13406	Potato solanidine
11	391	16.2	488	1 W84569	UDP-glucose:flavon
12	369.5	15.3	473	1 W36451	Anthocyanidin-3-gl
13	326.5	13.5	471	1 R49245	HUG-Br2. Isolated
14	145.5	6.0	534	1 R26154	HUG-Br1. Isolated
15	139.5	5.8	533	1 R26153	Ecdysteroid UDP-gl
16	121.5	5.0	506	1 R10429	Uridine diphospho-
17	121.5	5.0	530	1 W47126	Ecdysteroid UDP-gl
18	119	4.9	515	1 W6750	Saccharothrix aro
19	109	4.5	414	1 R3982	UGT1 Exon 4 produc
20	107	4.4	74	1 R30165	S. erythraea desosa
21	107	4.4	421	1 W99388	Glycosyltransferas
22	105.5	4.4	409	1 W71464	S. cerevisiae L947
23	101	4.2	473	1 W64392	S. cerevisiae L947
24	101	4.2	1198	1 W64384	Sugar biosynthesis
25	100	4.1	421	1 W19735	Zeaxanthin glycosy
26	98	4.0	399	1 R13989	Torenia flavonoid
27	94	3.9	508	1 W35712	Bispecific tetrava
28	93	3.8	580	1 W90217	Enterococcus faeca
29	88	3.6	431	1 Y00181	Enterococcus faeca
30	88	3.6	474	1 Y00180	Phosphofructokinase
31	87.5	3.6	781	1 W11298	Glucosyltransferase
32	87.5	3.6	1051	1 W52304	S. antibioticus des
33	87	3.6	426	1 W99397	

34	87	3.6	448	1 W64391	A. thaliana SGT pr
35	87	3.6	637	1 W64388	A. thaliana SGT pr
36	87	3.6	674	1 W64389	A. thaliana protei
37	87	3.6	1142	1 W50145	Mesotaenium caldar
38	87	3.6	1390	1 W21636	Grapevine leafroll
39	86.5	3.6	431	1 R07464	Polypeptide with e
40	86.5	3.6	431	1 W87890	Protein encoded by
41	85	3.5	547	1 R39741	ICAM-R (Intercellu
42	85	3.5	793	1 Y00869	S. tuberosum isoam
43	84.5	3.5	410	1 W66103	Human type 2 neuro
44	84.5	3.5	758	1 R23165	Mutant thermostabl
45	84.5	3.5	789	1 R23164	Mutant thermostabl

ALIGNMENTS

RESULT 1
W92947
ID W92947 standard; Protein: 460 AA.
AC W92947;
DT 14-MAY-1999 (first entry)
DE WO9905287 Seq ID 7.
KW Plant: flavonoid 5-transglycosylation activity; 5TG; variety; colour.
OS Perilla frutescens.
PN WO9905287-AL.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M.
DR WPI: 99-142940/12.
DR N-PSDB: X02826.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Claim 2; Page 71-73; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
CC
SQ Sequence 460 AA;

Query Match 100.0%; Score 2421; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 4.8e-248;
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVRRLVLLATFPAQGHINPALQFAKRLKAGCTDVTFTTSVYAWRRMANTASAAAGNPGL	60
Db	1	MVRRLVLLATFPAQGHINPALQFAKRLKAGCTDVTFTTSVYAWRRMANTASAAAGNPGL	60
Qy	61	DFVAFSDGYDGLKPCGDKRYMSEMKARGSEALNLLNNHVDVYVYSHLFAAAEVA	120
Db	61	DFVAFSDGYDGLKPCGDKRYMSEMKARGSEALNLLNNHVDVYVYSHLFAAAEVA	120
Qy	121	RESQVPSALLWVEPATVLCIYFYFYNGYADEIDAGSDEIQLPRLPLEQSLPTFLPET	180
Db	121	RESQVPSALLWVEPATVLCIYFYFYNGYADEIDAGSDEIQLPRLPLEQSLPTFLPET	180
Qy	181	PERFLRMKEKLTLDGEEKAKVLNTFDALPDALTATDRYELIGTIGPLPSAFLDGGD	240
Db	181	PERFLRMKEKLTLDGEEKAKVLNTFDALPDALTATDRYELIGTIGPLPSAFLDGGD	240
Qy	241	PSETSYGGDLFEKSENNCNEVWLDTPKPKSSVYVYFGSVLRFPPKAQMEIEGKGLACGRP	300
Db	241	PSETSYGGDLFEKSENNCNEVWLDTPKPKSSVYVYFGSVLRFPPKAQMEIEGKGLACGRP	300
Qy	301	FLWMIRKNDKDEEEEEEISCIIGELKKMGKIVSWCSOLEVLAHLPALGCFVTHCGWNSAV	360
Db	301	FLWMIRKNDKDEEEEEEISCIIGELKKMGKIVSWCSOLEVLAHLPALGCFVTHCGWNSAV	360

```
QY 361 ELSGCVPVVAVPQWFQDQTTNAKLIEDAWGTGVRVRNNEGGVDSGEIERCVERVMWDGGE 420
DB 361 ELSGCVPVVAVPQWFQDQTTNAKLIEDAWGTGVRVRNNEGGVDSGEIERCVERVMWDGGE 420

QY 421 KSKLVRENAIKWKTAREAMGEGDSSSLKLNAPFLHQAARA 460
DB 421 KSKLVRENAIKWKTAREAMGEGDSSSLKLNAPFLHQAARA 460

RESULT 2
W92948
ID W92948 standard; Protein; 443 AA.
AC W92948;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 8.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
OS Perilla frutescens.
PN W0905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR N-PSDB; X02827.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Claim 2; Page 74-76; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
SQ Sequence 443 AA;

Query Match 91.8%; Score 2223; DB 1; Length 443;
Best Local Similarity 95.7%; Pred. No. 4.4e-227;
Matches 424; Conservative 9; Mismatches 8; Indels 2; Gaps 1;

QY 1 WYRRVLLATPAQGHINPALQFAKRLKAGTDVFTTSYAWRRMANTASAAAGNPGL 60
DB 1 WYRRVLLATPAQGHINPALQFAKRLKAGTDVFTTSYAWRRMANTASAAAGNPGL 60

QY 61 DFVAFSDGYDDGLKPCGDKRYNEMKARGSEALRNLLNNHNDVTFVYSHLFANAEEVA 120
DB 61 DFVAFSDGYDDGLKPCGDKRYNEMKARGSEALRNLLNNHNDVTFVYSHLFANAEEVA 120

QY 121 RESQVPSALLWVEPATVLCIYFFYFNGYADEIDAGSDEIQLPRLPLEQORSPTLLPET 180
DB 121 RUSHVPTALLWVEPATVLCIYFFYFNGYADEIDAGSDEIQLPRLPLEQORSPTLLPET 180

QY 181 PERFRLLMKKELETLDGEEKAKVLTNTFDALPDALTAIDRVELIGIGLIPSAFLDGGD 240
DB 181 PERFRLLMKKELETLDGEEKAKVLTNTFDALPDALTAIDRVELIGIGLIPSAFLDGGD 240

QY 241 PSETSYGDLPEKSEENNCNVEMLDTPKSSVYVFGSVLRPPKQMEIIGKLLACGRP 300
DB 241 PSETSYGDLPEKSEENNCNVEMLDTPKSSVYVFGSVLRPPKQMEIIGKLLACGRP 300

QY 301 FLWMIREQKNDG--EEEBEELSCIGELKMKGIKIVSMCSQLEVLAPALGCFVTHCGWNS 358
DB 301 FLWMIREQKNDGEEEBEELSCIGELKMKGIKIVSMCSQLEVLAPALGCFVTHCGWNS 360

QY 359 AVESLSCGVPVAVPQWFQDQTTNAKLIEDAWGTGVRVRNNEGGVDSGEIERCVERVMWDG 418
DB 361 AVESLSCGVPVAVPQWFQDQTTNAKLIEDAWGTGVRVRNNEGGVDSGEIERCVERVMWDG 420

QY 419 GEKSKLVRENAIKWKTAREAMG 441
DB 421 GSKTKLVRENAIKWKTAREAMG 443
```

```
RESULT 3
W92949
ID W92949 standard; Protein; 461 AA.
AC W92949;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 9.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
OS Verbena hybrida.
PN W0905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR N-PSDB; X02828.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Claim 2; Page 76-78; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
SQ Sequence 461 AA;

Query Match 68.0%; Score 1646.5; DB 1; Length 461;
Best Local Similarity 70.0%; Pred. No. 6.2e-166;
Matches 326; Conservative 49; Mismatches 74; Indels 17; Gaps 7;

QY 1 WYRRVLLATPAQGHINPALQFAKRLKAGTDVFTTSYAWRRMANTASAAAGNPGL 60
DB 1 MSRAHVLLATPAQGHINPALQFAKRLANADIQVFTTSYAWRRMNTSRT---AAGSGLI 57

QY 61 DFVAFSDGYDDGLKPCGDKRYNEMKARGSEALRN--LLNNHND----VTFWYSHLFA 114
DB 58 NFVSFSDGYDDGLQPGDDGKNYSEMSKRGKALSDDTLANNVDDQKSKITPVYSHLFA 117

QY 115 WAAEVARESQVPSALLWVEPATVLCIYFFYFNGYADEIDAGSDEIQLP-RLPLEQORSLP 173
DB 118 WAAKVARFELHRSALLWTEPATVLDIIFYFNGYSDIAGSDAIHLPGGLVLAQROLP 177

QY 174 TELLPETFRFLMKKELETLDGEEKAKVLTNTFDALPDALTAIDRVELIGIGLIPS 233
DB 178 SELLPTHERFRSLMKKELETLEGEKPKVLVNSFDALPDALKADKYEMTAIGLIPS 237

QY 234 AFLDGGDSETSYGDLPEK--SEENNCVENLDTKPKSSVYVFGSVLRFPKQMEIEIGK 292
DB 238 AFLDGGDPSDFSFGDULFEKGSNDCCLEWLSTNPKSSVYVFGSVFVNTTKSQMEIEAR 297

QY 293 GLLACGRPLFWMIREQKNDGEEEBEELSCIGELKMKGIKIVSMCSQLEVLAPALGCFVT 352
DB 298 GLLDGRPLFWMVRV----NEGDEVLTSCMEELKRVKIVSMCSQLEVLTHPSLGCFVT 352

QY 353 HCGWNSAVESLSCGVPVAVPQWFQDQTTNAKLIEDAWGTGVRVRNNEGG--VDGSEIERC 411
DB 353 HCGWNSTLESISFGVPVAVPQWFQDQTNKALMEDVWRTGVRVRANEESVVDGDEIRRC 412

QY 412 VEMVMDGGEKSKLVRENAIKWKTAREAMGEGDSSSLKLNAPFLHQAARA 457
DB 413 IEEVMDGGEKSKLRRESAGKWKOLARKAMEEDGSSVNNLKVFLDEV 458

RESULT 4
W92950
ID W92950 standard; Protein; 478 AA.
AC W92950;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 10.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
```


PT Plants genetically transformed to interfere with
 PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression
 PT - useful for production of rapeseed oil with reduced glucosinolate
 PT content
 PS Claim 9; Page 23-25; 35pp; English.
 CC A UDP-glucose:thiohydroximate S-glucosyltransferase (S-GT) (W09825)
 CC is encoded by clone pGU9 (T66166) amplified from *Brassica napus* cv.
 CC Westar cDNA. S-GT is the enzyme responsible for the biosynthesis
 CC of glucosinolate. Novel chimeric genes encode an antisense RNA
 CC complementary to all or part of an mRNA, the cDNA of which is
 CC contained in pGU9. Oilseed rape plants transformed with these
 CC chimeric genes have reduced contents of glucosinolates, pref.
 CC alkenyl glucosinolates. This allows the prodn. of rapeseed oil
 CC with a low glucosinolate content.
 SQ Sequence 466 AA;

Query Match 28.94; Score 700; DB 1; Length 466;
 Best Local Similarity 35.7%; Pred. No. 1.4e-65;
 Matches 169; Conservative 75; Mismatches 185; Indels 44; Gaps 13;

QY 6 VLLATFPAGQHINPALQAKRLKAGTDVTFSTVYAMRMAMTASAAAGNPPGLDFAV 65
 DB 17 VVLPVPVQGLHNPVQAKRLKGVKVTIATTY-----TASSISTPSVSEPI 67
 QY 66 SDGYDDGL--KPCGDKKRYMSEMKARGSEALRNLL-----LNNHDVTFVYSHLFAWAEV 119
 DB 68 SDGHDPILGVPVCSIDAYSEFKLNGSETLTVIRISFKSTSPIDSLVYDSFLPWGLEV 127
 QY 120 ARESQVPSALLWVEPATVLCIYFFNGYADEIDAGSDEIQLP---RLPPLQORSLPTF 175
 DB 128 ARSNSISAAFTNNLTWCISVLRNFVSG---EFLPADPASAPLYRGLPALSYDELPSF 184
 QY 176 L----LPETPERFLMMKEFLTOGEKAKVLYNTFPALEPDALTAIDRYELIG--IGPL 230
 DB 185 VGRHSSHAHGRVLLN---QFRNHEDADWLFVNSFGLTQCEVGESEAMRATLIGPM 241
 QY 231 IPSAFIDGGDPSYSGDGLPEKSENNCNVWLDTPKSSVYVVSFGSVLRFPKQMEET 290
 DB 242 IPSAYLDGRDKDKGSSLMKPLSE-ECNEWLDTKLSKVVSFSGFGLPFEKQLAEV 300
 QY 291 GKLLACGPRFLWNTRE--QKNDGDEEEELSCIGELKMKKIYVSWCSQLEVLAPAL 347
 DB 301 AKALQESNFNLWVKEAHIAKLPEGFVE-----ATKDRALLVSWCNQLEVLAVHSI 352
 QY 348 GCFVTHCGWNSAVESLSCGVPVAVPWQFQDTTNAKLIEDAWGTGVRVRNDEGG-VDGS 406
 DB 353 GCFLTHCGWNSLLEGLSLGVPVGPQWSQMDNDKAFVEVWRVGRYRAKEAGGVKSD 412
 QY 407 EIERCVENWDGKSKLVRENATKWTLAREANGDCGSLKKNLAPLHOVAR 459
 DB 413 EYVRCILRGYME-GESSVSEIRSSKKWKDLAYKAMSEGGSSDRSINEFVESLKG 464

RESULT 8
 W98009
 ID W98009 standard; Protein; 471 AA.
 AC W98009;
 DT 21-JUN-1999 (first entry)
 DE Medium-chain UDP-glucose:fatty acid glucosyltransferase.
 KW UDP-glucose:fatty acid glucosyltransferase; tomato;
 KW 1-O-acyl-a-glucose; acyl donor; esterification.
 OS Lycopersicon pennellii
 OS Lycopersicon esculentum.
 PN W0909144-AL
 PD 23-FEB-1999.
 PF 10-AUG-1998; IB1369.
 PR 30-JUN-1998; US-106464.
 PR 13-AUG-1997; US-055554.
 PA (CORR.) CORNELL RES FOUND INC.
 PI Eanneatta NT, Ghangas GS, Kuai J, Steffens JC;
 DR WPI: 99-190155/16.
 DR N-PSDB; X24873.

PT New chain length specific uridine diphosphate-glucose:fatty acid
 PT glucosyltransferases - useful for preparing straight or branched,
 PT (unsaturated) 1-O-acyl-a-glucoses and 1-O-acyl-a-D-glucoses
 PS Example 3; Page 36-38; 42pp; English.
 CC The present sequence is a UDP-glucose:fatty acid glucosyltransferase
 CC that transfers glucose to medium-chain length fatty acids to form
 CC 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses act as acyl donors
 CC in the esterification of glucose and in the further esterification
 CC of partially acylated glucose in reactions catalysed by glucose
 CC acyltransferases. They also act as acyl donors in the
 CC esterification of sucrose and in the further esterification of
 CC partially acylated sucrose in reactions catalysed by sucrose
 CC acyltransferases. A cDNA clone (see X24873) encoding the enzyme
 CC was isolated by immunoscreening a leaf trichome cDNA library
 CC prepared from an F1 population of a cross between wild tomato
 CC Lycopersicon pennellii (LA 716) and cultivated tomato
 CC Lycopersicon esculentum cv. New Yorker. The invention provides
 CC chain length-specific UDP-glucose:fatty acid glucosyltransferases
 CC and a method for their purification. The enzymes have specificity
 CC for transferring glucose to short, medium or long chain length
 CC fatty acids. Methods for preparing 1-O-acyl-a-D-glucoses, where
 CC the acyl group comprises 3-5, 6-13 or 14-22 C atoms, are claimed.
 SQ Sequence 471 AA;

Query Match 16.8%; Score 407.5; DB 1; Length 471;
 Best Local Similarity 28.3%; Pred. No. 1.5e-34;
 Matches 140; Conservative 66; Mismatches 182; Indels 107; Gaps 20;

QY 11 FP--AOGHNPALQAKRLKAGTDVTFST---SYVAMRRMAN-----TAS 51
 DB 10 FPMAGCHMPTLDMAKLVACRGVKATIIITPLNESVFSKAIERNKHLGIEIDIRLLKPP 69
 QY 52 AAAGNPP----GLDFVAFSDGYDDGLKPCGDKKRYMSEMKARGSEALRNLLNHDVTFV 107
 DB 70 AKENDLPEDCELDLVPDDKLPNFKAAAMKDEFE-----LIGECRPDCL 117
 QY 108 VYSHLFAWAAEVARQSPVPSALLWVEPATVLCIYVYFNGYADEIDAGSDEIQLPRLPPL 167
 DB 118 VSDMFLPTWTDNAKESIPRIVFHGTSYFALCVGH-----SIRNKKPF 160
 QY 168 EQRS--LPTFLPEPERRPLMKKE--KLETL-----GEEKAK---VLVNTF 208
 DB 161 KNYSSDTFTFVDPDPHEIRITRTQLSPFQSDDEETGMAPIKAVRESKASGYVFNFSF 220
 QY 209 DALEPDALTAIDRY-ELIG-----IGPLPSAFIDGGDPSYSGDLEFKS--EENNCV 260
 DB 221 YELED---YVEHYTKVGRKNWAIGPLSLC-----NRDIEYKAERGRSSIDEHACL 270
 QY 261 EWLDTPKSSVYVVSFGSVLRFPKQMBEIGKGLACGRPFLWIRQKND---DGEEEE 317
 DB 271 KWLDSKSSSIYVVCVGSFADFTTAQMQLAMGLEASGQDFIWIWRTGDNEDWLPKGFEE- 329
 QY 318 EELSCIGELKMKKIV-SWCSOLEVLHAPALGCFVTHCGWNSAVESLSCGVPVAVPWQF 376
 DB 330 -----RTKEKGLIRGWAPQVLIDHEAGAFVTHCGWNSLLEGISAGVPMITWVPFA 382
 QY 377 DQTTNAKLIEDAWGTGVRV-----RNNEGGVVDGSEIERCVEMVMDGKSKLVRENATK 431
 DB 383 EOPFNEKLVTVMRSAGVSGKQWKRTASEGVKREAIKAKIKRVN-ASETEGFRSRAKE 441
 QY 432 WKTLAREANGEDGSS 446
 DB 442 YKEMAREATIEGGSS 456

RESULT 9
 W47172
 ID W47172 standard; Protein; 470 AA.
 AC W47172;
 DT 08-JUN-1998 (first entry)
 DE Glucosyl transferase (Glase) protein encoded by TWII gene.
 KW Glucosyl transferase; Glase; TWII; tomato; signalling pathway;

KW salicylic acid; jasmonic acid; ethylene; wound inducible gene;
 OS plant defence protein; plant response; tobacco; rice.
 SW Lycopersicon sp.
 PN W09745546-A1.
 PD 04-DEC-1997.
 PF 30-MAY-1997; G01473.
 PP 31-MAY-1996; GB-011420.
 PA (UYVO-) UNIV YORK.
 PI Bowles DJ, Calvert CM, Odonnell PJ, Roberts MR;
 DR WPI; 98-032653/03.
 DR N-PSDB; V17054.
 PT tomato wound inducible (TWI1) gene encoding glucosyl transferase -
 PT useful to develop products that alter signalling pathways in plants
 PT by altering of salicylic acid, jasmonic acid or ethylene
 PS Claim 2; Fig 3; 52pp; English.
 CC This is a glucosyl transferase (Grase) protein encoded by a wound
 CC inducible gene (TWI1) isolated from wounded tomatoes. The TWI1 gene
 CC encodes this Grase from amino acid position 5. The TWI1 gene can be
 CC used to identify homologue Grase encoding genes isolated from tobacco
 CC and rice. A microbial host can be transfected or transformed with a
 CC vector containing the Grase encoding nucleic acids. The products can be
 CC used to interfere with Grase and therefore alter signalling pathways in
 CC plants, specifically tobacco, rice or tomato plants by altering levels
 CC of salicylic acid, jasmonic acid or ethylene. This can induce the
 CC production of plant defence proteins such as pathogenesis-related (PR)
 CC and proteinase inhibitor (PIN) proteins which regulate plant development
 CC (plant growth, reproduction and senescence) and improve plant response to
 CC pathogens.
 SQ Sequence 470 AA;

Query Match 16.8%; Score 406.5; DB 1; Length 470;
 Best Local Similarity 28.3%; Pred. No. 1.9e-34;
 Matches 142; Conservative 64; Mismatches 176; Indels 119; Gaps 21;

QY 11 FP--AOGHINPALQFAKRLKAGTDVTF--SVYAWRMAN-----TAS 51
 DB 9 FPDDAOGHMTPLDMANVACRGVKAITITPLNESVFSKATERNKHILGIDIRLLKFP 68
 QY 52 AAAGNPP---GLDEFAVSQDGLKPCGDKGRYMSKARGSEALRNLLNHDVTFV 107
 DB 69 AKENDLPEDCERLDVPSDDKLPNFKAAAMKDFEE-----LIGCRPDCL 116
 QY 108 VSHLFWAAEVAESQVPSALLWEPATVLCIYFYFNGYADEIDAGSDIQLPLPL 167
 DB 117 VSDMFLPMTDSAAKFSIPRIVPHGTSYFALCV-----GDTIR--RNKPF 159
 QY 168 BORS--LPTFLLPETPERFLMKKE--KLETLD-----GEEKAK---VLVNTF 208
 DB 160 KNVSSDTETFWPDLPHREIRLTQLSPEQSDDEETGNAPMIKAYRESDAKSYGVINSP 219
 QY 209 DALPEDALTAIDRY-ELIG-----IGPLIPSAFLDGGDPSYSGDLPKSE----- 255
 DB 220 YELES---YVEHYTKVYGRKNNAIGPL-----SLCRNRDIEDKAERGKSSI 263
 QY 256 -ENNCVEMLDTPKSSVYVFGSVLFRPRAQMEIEIGKGLLACGRPFLWMIREQND--- 311
 DB 264 DEHACLKWLDSKSSIVYVFCGSTADFTTAQMOELAMGLASGQDFIWTIRGNEWLP 323
 QY 312 DGEEBEELSCIGELKNGKIV-SWCSQLEVLAHPALGCFVTHCGNSAVESLSCGVPV 370
 DB 324 EGFE-----RTKEKGLIIRGWAQPSVILDHGAIFVTHCGNSITLEGISAGVPMV 375
 QY 371 AVPQWFDQTTNAKLIEDAWGTGRV-----RMNEGVDGSGSERCVEMVMDGGEKSLV 425
 DB 376 TWPVFAEQFTNEKLVTEVMRSGAGVSKQWKRTASGVKREIAKIRVM-ASEETEGF 434
 QY 426 RENAUKWTLAREANGDGSS 446
 DB 435 RSRAKEYKEMAREAIPEGSS 455

RESULT 10

WL3406
 ID W13406 standard; Protein; 433 AA.
 AC W13406;
 DT 10-JUN-1997 (first entry)
 DE Solanum melongena flavanoid-3-glucosyl transferase.
 KW Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;
 OS Solanum melongena.
 PN J09053385-A.
 PD 04-MAR-1997.
 PF 25-AUG-1995; 238943.
 PP 25-AUG-1995; JP-238943.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI; 97-206631/19.
 DR N-PSDB; T62123.
 PT Solanum melongena flavanoid-3-glucosyl transferase and DNA encoding
 PT it - catalyses transfer of glucose to 3-position of flavonoid
 PT pigment in plants
 PS Claim 4; Pages 9-10; 14pp; Japanese.
 CC The present sequence is the Solanum melongena (eggplant)
 CC flavanoid-3-glucosyl transferase (F3GT), which catalyses the
 CC transfer of glucose to the 3-position of flavonoid pigments. The
 CC F3GT DNA is useful for the production of the seed of a plant. A
 CC S. melongena seed was cultured under a red light and a flavonoid
 CC derived. A cDNA library in which the pigment was concentrated was
 CC prepared. A flavonoid glucosylidising enzyme gene was cloned and its
 CC structure and mRNA expression pattern analysed. The clone was
 CC expressed in E. coli.
 SQ Sequence 433 AA;

Query Match 16.4%; Score 398; DB 1; Length 433;
 Best Local Similarity 26.7%; Pred. No. 1.3e-33;
 Matches 124; Conservative 87; Mismatches 176; Indels 78; Gaps 20;

QY 6 VLLATFFPAQGHINPALQFAKRL---LKAGTDVTF--SVYAWRMANTASAAAGNPPGLDF 62
 DB 8 IATLAFPGTGTATPLTLVQKISPLSPSTIFSFT-----SSNSSIFSKVPNDEN 60
 QY 63 VAFSGDYDGLKPCGDKGRYMSKARGSEALR-----NLLN--HDTFTV---V 108
 DB 61 KIYNW-DGVKEND-----TPGLEAIKLFQSTLLISKITEAEETGVKFSKI 111
 QY 109 YSHLFWAA--AEVARESQVPSALLWEPATVLCIYFYFNGYADEIDAGSD-EIQLPL- 164
 DB 112 FSDAFLMCFVLKLPKKNAPGAVYWTGGCSLAVHL-----YTDLIRSKETSLKPGFS 166
 QY 165 PLEQRSPLTFLPE-TPERFLMKKEKL--ETLDGEEKAKVLVNTFDALPDALTAID- 220
 DB 167 STLINDIP----PEVTAEDLEGPMSSMLYNMALNHLKADAVVLNSFQELDRDPLINKDL 222
 QY 221 ---RYELIGIPLPSAFLDGGDPSYSGDLPKSENNCVEMLDTPKRSVYVYVFG 277
 DB 223 QKNLQKVNIGPLVQS-----SRKLDSEGCIOWLDKQKESVYVLSFG 266
 QY 278 SVLRPRAQMEIEIGKGLLACGRPFLWMIREQNDGEEBEELSCIGELKNGKIVSWCS 337
 DB 267 IVTTLPPNEIGSIAEALETKTTPINSR----NNGVKNLPK-GFLERTKEFGKIVSNAP 321
 QY 338 QLEVLAHPALGCFVTHCGNSAVESLSCGVPVAVPQWFDQTTNAKLIEDAWGTGRV 397
 DB 322 QLEILAHKSVGFVTHCGNSITLEGISFGVPMICRPFPGDQKLNRSWVESVMEIGLIQI-- 379
 QY 398 NEGGVDGSGSERCVEMVMDGGEKSKLVRENAIKWTLAREANGE 442
 DB 380 EGGIFTSGSIISALDFFN-BEKGKILRENVEGLEKALEAVNQ 422

RESULT 11
 W64569
 ID W64569 standard; Protein; 488 AA.
 AC W64569;
 DT 07-DEC-1998 (first entry)


```
OS Homo sapiens.
FH Key Location/Qualifiers
FT 10..20
FT /note= "putative membrane-insertion signal"
FT 491..507
FT /note= "putative membrane-anchoring peptide"
FT modified_site 102
FT /note= "predicted Asn-linked glycosylation site"
FT modified_site 295
FT /note= "predicted Asn-linked glycosylation site"
FT modified_site 347
FT /note= "predicted Asn-linked glycosylation site"
FT misc_difference 158
FT /note= "feature not labelled in specification"
FT misc_difference 181
FT /note= "feature not labelled in specification"
FT misc_difference 228
FT /note= "feature not labelled in specification"
FT WO9212987-A.
PN 06-AUG-1992.
PD 10-JAN-1992; U00282.
PR 10-JAN-1991; US-639453.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Owens IS, Ritter JK;
DR WPI: 92-284593/34.
DR N-PSDB: Q27369.
PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT types I and II
PS Disclosure; Fig 9A-I; 99pp; English.
CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC been isolated. They are referred to as HUGBr1 (Q27369) and HUGBr2
CC (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC upon expression individually in COS-1 cells, encode isoforms that
CC catalyse the formation of the two bilirubin monoglucuronides and
CC the diglucuronide.
CC The cDNAs contain identical 3' ends (1469 bp in length) to each
CC other and to that of the human phenol transferase cDNA, HUGP1
CC (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC In contrast, they have unique 5' ends.
SQ Sequence 533 AA;

Query Match 5.88; Score 139.5; DB 1; Length 533;
Best Local Similarity 22.28; Pred. No. 4.7e-06;
Matches 49; Conservative 38; Mismatches 85; Indels 49; Gaps 7;

QY 229 PLIPSAFLDGGPPSETSYGGDLFEKSENNCVE-----WLDTPKSSVYVYSGS 278
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 267 PIMPNNVFGG-----INLHONPLSQEFAYINASGEHGIVVFSLES 309
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 279 -VLRFPKQAMEIGKGLACGRPFIMIRBOKNDGDGEEELSCITGELKKMGKIVSWCS 337
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 MYSEIPEKKAMAIAADALGKIPQTVLWRYTCTRPSN-----LANNTILYKWL 356
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 338 QLEVLHAPALGCFVTHCGWNSAVESLSGVPVYAVPOWFDQTTNAKLIEDAWGTGVVRM 397
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 357 QNDLLCHPMTAFITHAGSHGVYESICNGVPVWVWMLFGQMDNARMETK-GAGVTLV 415
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 398 NEGGGVDSIEIRCEVMDGGEKSLVRENAIKWTKLARE 438
|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 416 LE---MTSEDLNQAQAVIN---DKSYKENIMRLSLHDK 449
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 1, 2000, 14:38:58 ; Search time 59,57 seconds
(without alignments)
118,372 Million cell updates/sec
Title: US-09-147-955-2
Perfect score: 2421
Sequence: 1 MVRRLVLLATPPAOGHINP.....GEGSSILKLNALHGVARA 460
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 152396 seqs, 15329161 residues
Total number of hits satisfying chosen parameters: 152396
Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description
1 762.5 31.5 471 2 US-08-466-583-2 Sequence 2, Appl
2 762.5 31.5 471 4 PCT-US95-07820-2 Sequence 2, Appl
3 409.5 16.9 471 3 US-09-106-464-2 Sequence 2, Appl
4 394.5 16.3 131 2 US-08-466-583-5 Sequence 5, Appl
5 394.5 16.3 131 4 PCT-US95-07820-5 Sequence 5, Appl
6 391 16.2 488 2 US-08-797-226-2 Sequence 2, Appl
7 245.5 10.1 131 2 US-08-466-583-6 Sequence 6, Appl
8 245.5 10.1 131 4 PCT-US95-07820-6 Sequence 6, Appl
9 220 9.1 63 2 US-08-466-583-8 Sequence 8, Appl
10 220 9.1 63 4 PCT-US95-07820-8 Sequence 8, Appl
11 166.5 6.9 529 4 PCT-US92-00282-7 Sequence 7, Appl
12 160.5 6.6 531 4 PCT-US92-00282-6 Sequence 6, Appl
13 153.5 6.3 534 4 PCT-US92-00282-4 Sequence 4, Appl
14 148.5 6.1 533 4 PCT-US92-00282-5 Sequence 5, Appl
15 148.5 6.1 533 4 PCT-US92-00282-3 Sequence 3, Appl
16 121.5 5.0 506 5 5180581-2 Patent No. 5180581
17 120 5.0 56 2 US-08-466-583-4 Sequence 4, Appl
18 120 5.0 56 4 PCT-US95-07820-4 Sequence 4, Appl
19 109 4.5 414 2 US-08-750-524-1 Sequence 1, Appl
20 107 4.4 74 4 PCT-US92-00282-24 Sequence 24, Appl
21 106 4.4 58 2 US-08-466-583-9 Sequence 9, Appl
22 106 4.4 58 4 PCT-US95-07820-9 Sequence 9, Appl
23 105.5 4.4 409 2 US-08-924-254-2 Sequence 2, Appl
24 105.5 4.4 409 3 US-09-120-249-2 Sequence 2, Appl
25 102 4.2 52 2 US-08-466-583-7 Sequence 7, Appl
26 102 4.2 52 4 PCT-US95-07820-7 Sequence 7, Appl
27 100.5 4.2 1114 2 US-08-576-626A-31 Sequence 31, Appl
28 100 4.1 421 2 US-08-576-626A-53 Sequence 53, Appl

ALIGNMENTS
RESULT 1
US-08-466-583-2 Application US/08466583
; Sequence 2, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jedrej B.
; APPLICANT: Szygowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08466.583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265.427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-583-2
Query Match 31.5%; Score 762.5; DB 2; Length 471;
Best Local Similarity 39.1%; Pred. No. 1.3e-71;
Matches 187; Conservative 70; Mismatches 182; Indels 39; Gaps 16;

```

QY 6 VLLATFPAQGHINPAQFAKRLKAGTDTFTTSYAHRRMANTASAAAGNPPGLDFVAF 65
DB 5 VLVVPPFGQGHMNPVQFAKRLKAGTDTFTTSYAHRRMANTASAAAGNPPGLDFVAF 57
QY 66 SDGYDDGLKPCGCG-KRYASEMKARGSEALRNL---LNNHDV-TFVVYSHLFAAAEVA 120
DB 58 SDGHEGFFASAGVAEYLEKQAAASASLASLVEARASSADAFCTCVYDSDYEDWVLPVA 117
QY 121 RESQVPSALLWEPATVLCIYFYFNG-----YADEIDAGSDEIQLP-----LPPL 168
DB 118 RMGLPAVPFSTQSCAVSAVYHFSGRLAVPPGAADGSDGAGAAALSEAFGLPEME 177
QY 169 QRSUPTLLPETPERPLMKKELETLDGEKAKVLVNTFDALPDALTADRY-ELIGI 227
DB 178 RSELPFVDFHGPYPTIAMQAIQFAHAGKDD-WVLFNSFELETEVLGLTKYLKARAI 236
QY 228 GPLI--PSAFLDGDGDPSETSYGGDLPEKSEENNCVEMWLDTPKPSVYVSGSVLRPPKA 285
DB 237 GPCVPLPTAGTAGANGRITYGANLVK--PEDACTKWLDTKPDRAVYVSGSVLASLGN 294
QY 286 QMEIGKGLLAGCPFLMWIREQKNDGDEEEELSCIGELKMG--KIVSWCSQLEVLA 343
DB 295 QKEELARGLLAAGKPFLLVYRA-----SDEHQPVRYLLAEATATGAAMVVPWCPQLDVLA 349
QY 344 HPALGCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTNAKLIEDWGTGVRVRNNEGGV 403
DB 350 HPVGCFCVTHCGWNSTLEALSFGVPMVAMALWTDQPTNARNVELAWGAGVARRDAGV 409
QY 404 D-GSEIERCVMVMDGGEKSKLVRENAIKWKTILAREAMGDEGSSLLKNLNAFLHQA 460
DB 410 FLRGEVRCVRAVMDGGEAASAKAAGEWRDRARAAPVAGSSDRNLDEEV-QFVRA 466

```

RESULT 2

```

PCT-US95-07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszyn, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07820-2

Query Match 31.5%; Score 762.5; DB 4; Length 471;
Best Local Similarity 39.1%; Pred. No. 1.3e-71;
Matches 187; Conservative 70; Mismatches 182; Indels 39; Gaps 16;

QY 6 VLLATFPAQGHINPAQFAKRLKAGTDTFTTSYAHRRMANTASAAAGNPPGLDFVAF 65
DB 5 VLVVPPFGQGHMNPVQFAKRLKAGTDTFTTSYAHRRMANTASAAAGNPPGLDFVAF 57
QY 66 SDGYDDGLKPCGCG-KRYASEMKARGSEALRNL---LNNHDV-TFVVYSHLFAAAEVA 120
DB 58 SDGHEGFFASAGVAEYLEKQAAASASLASLVEARASSADAFCTCVYDSDYEDWVLPVA 117
QY 121 RESQVPSALLWEPATVLCIYFYFNG-----YADEIDAGSDEIQLP-----LPPL 168
DB 118 RMGLPAVPFSTQSCAVSAVYHFSGRLAVPPGAADGSDGAGAAALSEAFGLPEME 177
QY 169 QRSUPTLLPETPERPLMKKELETLDGEKAKVLVNTFDALPDALTADRY-ELIGI 227
DB 178 RSELPFVDFHGPYPTIAMQAIQFAHAGKDD-WVLFNSFELETEVLGLTKYLKARAI 236
QY 228 GPLI--PSAFLDGDGDPSETSYGGDLPEKSEENNCVEMWLDTPKPSVYVSGSVLRPPKA 285
DB 237 GPCVPLPTAGTAGANGRITYGANLVK--PEDACTKWLDTKPDRAVYVSGSVLASLGN 294
QY 286 QMEIGKGLLAGCPFLMWIREQKNDGDEEEELSCIGELKMG--KIVSWCSQLEVLA 343
DB 295 QKEELARGLLAAGKPFLLVYRA-----SDEHQPVRYLLAEATATGAAMVVPWCPQLDVLA 349
QY 344 HPALGCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTNAKLIEDWGTGVRVRNNEGGV 403
DB 350 HPVGCFCVTHCGWNSTLEALSFGVPMVAMALWTDQPTNARNVELAWGAGVARRDAGV 409
QY 404 D-GSEIERCVMVMDGGEKSKLVRENAIKWKTILAREAMGDEGSSLLKNLNAFLHQA 460
DB 410 FLRGEVRCVRAVMDGGEAASAKAAGEWRDRARAAPVAGSSDRNLDEEV-QFVRA 466

RESULT 3
US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 6011145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Changas, Curdev S.
; APPLICANT: Kual, Jian-Ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc:Fatty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,464
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,554

```

```

, , FILING DATE: 13-AUG-1997
, , ATTORNEY/AGENT INFORMATION:
, , NAME: SPECTOR, ERIC S.
, , REGISTRATION NUMBER: 22495
, , TELECOMMUNICATION INFORMATION:
, , TELEPHONE: 703-415-1500
, , TELEFAX: 703-415-1508
, , INFORMATION FOR SEQ ID NO: 2:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 471 amino acids
, , TYPE: amino acid
, , TOPOLOGY: linear
, , MOLECULE TYPE: protein
, , US-97-106-464-2

```

[illegible]

```

RESULT      4
US-08-466-583-5
; Sequence 5, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
;

```

```

? COUNTRY: United States of America
? ZIP: 80303
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/456,583
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 800
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/265,427
? FILING DATE: 24-JUN-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Ferber, Donna M.
? REGISTRATION NUMBER: 33,878
? REFERENCE/DOCKET NUMBER: 11-94A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 303/499-8080
? TELEFAX: 303/499-8089
? TELEX: 49617824
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 131 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? HYPOTHETICAL: NO
? FRAGMENT TYPE: Internal
? US-08-456-583--5

```

Query Match 16.3%; Score 394.5; DB 2; Length 131;
Best Local Similarity 53.7%; Pred. NO. 7.6e-34;
Matches 73; Conservative 19; Mismatches 37; Indels 7

Qy	259	CVENLDTKPKSSVYVSGSVLRPPKQAEETGKGLACGPPFLWMLREQKNDGDEEEE	318
Db	1	CTKWLDTKPRVAVYSGSVLSLGNQAEELARGLLAKGPFLLWVRA-----SDEHQV	55
Qy	319	ELSCIGELKKMG--KIVSNCSQLEVLZLAHPALGCFVTHCGWNSAVESLSCGVPVYAVPQPF	376
Db	56	PRYLTAEAATATGAAMVPMWPCQLDVLHAHPVCGFVTHCGWNSLEALSFSGVPMYMALMT	115
Qy	377	DOTTNAKLIEDAWGTG	392
Db	116	DOTPNARNVELANGAG	131

```

RESULT 5
PCT-US95-07820-5
: Sequence 5, Application PC/TUS9507820
: GENERAL INFORMATION:
: APPLICANT: Bandurski, Robert S.
: APPLICANT: Szerzen, Jedrzej B.
: APPLICANT: Szczygłowski, Krzysztof
: TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
: TITLE OF INVENTION: and Plant Growth.
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee & Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/07820
;; FILING DATE: 19-JUN-1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/265,427
;; FILING DATE: 24-JUN-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Feiber, Dorra M.
;; REGISTRATION NUMBER: 33,878
;; REFERENCE/DOCKET NUMBER: 11-94B PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 303/499-8080
;; TELEFAX: 303/499-8089
;; TELEX: 49617824
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 131 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; FRAGMENT TYPE: internal
;; PCT-US95-07820-5

Query Match 16.3%; Score 394.5; DB 4; Length 131;
Best Local Similarity 53.7%; Pred. No. 7.6e-34;
Matches 73; Conservative 19; Mismatches 37; Indels 7; Gaps 2;

Qy 259 CVELWDTKPKSSVYVYVSGSVLRFPKAQMEEIGKGLACGPRFLWMIREQKNDGDEEE 318
Db 1 L-||||| || ||||| || ||: ||||| L-|||||: ||
Qy 319 ELSCEIGELKMG--KIVSWCSQLEVLHAPALGCFVTHCGWNSAVESLSCGVVPVAVPQWF 376
Db 56 PRYLAEATATGAAMVVPWCPQLDVLHAPVGCFTVHCGWNSTLEALSFGVPPVYAMALWT 115

Qy 377 DOTTNAKLIEDANGT 392
Db 116 DQPTNARNVELANGAG 131

RESULT 6
US-08-797-226-2
; Sequence 2, Application US/08797226
; Patent No. 5959180
; GENERAL INFORMATION:
; APPLICANT: MOEHS, CHARLES P
; APPLICANT: ALLEN, PAUL V
; APPLICANT: ROCKHOLD, DAVID R
; APPLICANT: STAPLETON, ANDREW
; APPLICANT: GARBARINO, JOAN E
; APPLICANT: FRIEDMAN, MENDEL
; APPLICANT: BELKNAP, WILLIAM R
; TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
; TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
; TITLE OF INVENTION: GLYCOALKALOIDS IN SOLANACEOUS PLANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NANCY J. PARSONS
; STREET: 800 BUCHANAN ST.
; CITY: ALBANY
; STATE: CA
; COUNTRY: USA
; ZIP: 94710
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/797,226
;; FILING DATE:
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: PARSONS, NANCY J
;; REGISTRATION NUMBER: 40,364
;; REFERENCE/DOCKET NUMBER: 0011.97
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 559-5731
;; TELEFAX: (510) 559-5777
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 488 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-797-226-2

Query Match 16.2%; Score 391; DB 2; Length 488;
Best Local Similarity 23.9%; Pred. No. 1.5e-32;
Matches 125; Conservative 89; Mismatches 186; Indels 122; Gaps 20;

Qy 6 VLLATFPAQGHINPALQFAKRLKACGDTFTFTSVYVARRNANT----- 49
Db 13 VLFPLFSAGHFIPLVNNAARLFAKRVKATILITPHNALLFRSTIDDDVRISGPIIVT 72

Qy 50 ---ASAAAGNPGLDFVAFSDGDDGLKPCGDKRY--MSEMKARGSEALNLLNNHDV 104
Db 73 IKFPSAEVGLPEGI-----ESPNSATSPMPHKIFAYALSILQKPMEDIKRELDPD--- 122

Qy 105 TFVYVSHL-FAWAAEAVARESQVPSALLWVEPATVLCIYFFV-FNCYADEIDAGSDEIQLP 162
Db 123 --CFSDMYFPTVADIADLHPIRL--YNLSAYWCYSIMINLKYVRPHKQPNLDESQ-- 176

Qy 163 RLPLEQORSPTLLPETPERFKLMKKELETL----- 204
Db 177 -----SFVPGPLPDEIKFKLSQTLDDLKRSDDQKTVFDELLEQVEDSEERSYGI 225

Qy 205 V-NTFDALPDALTADRYELI-----GIGPLIPSAFLDGGDPSETSYGDLFEK---S 254
Db 226 VHDTFEYLEP---AYVDYQKAKKPKCWFGFL-----SHFASKIRSKELIS 269

Qy 255 EENN---CVELWDTKPKSSVYVYVSGSVLRFPKAQMEEIGKGLACGPRFLWMIREQKND 311
Db 270 EHNNEIVIDWLNAAQPKSVLYVYVSGSMARFPESQLETAQALDASNVYFFIVLRP--- 325

Qy 312 DGEDEEEELSCIGELKMGK---IVSWCSQLEVLHAPALGCFVTHCGWNSAVESLSCGV 367
Db 326 --NEETASWLPVGNLEDKTKGLYIKGNVPQLTIMESHSATGCFVTHCGTNSVLAITFGV 383

Qy 368 PVAVPQWFQDTTNAKLIE-----DANGTGVYVRVMEGGVGDGSEIERCIVE--MV 415
Db 384 PMITWPLYADQFYNEKYVEVYVGLGIRIGIDVWNEGIEI---TGPVIESAKIREAIERLMI 440

Qy 416 MDGEGSKLVRENAIKWKTILAREAMGEDGSSLKLNLAFLHQV 457
Db 441 SNGSEELINIDRVMAWSKMAQNAATNEGGSSNNLTALIQHI 482

RESULT 7
US-08-466-583-6
; Sequence 6, Application US/08466583
; Patent No. 591998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jędrzej B.
; APPLICANT: Szczylowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.


```

; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; US-08-466-583-8

Query Match          9.1%; Score 220; DB 2; Length 63;
Best Local Similarity 67.3%; Pred. No. 5.1e-16;
Matches 37; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 332 IVSWCSQLEVLHAHPALGCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTNAKLIE 386
Db 9 VVPWCPQLDVLHAHPAVGCFVTHCGWNSTLEALSFGVPMVAMALWTDQPTNARNE 63

RESULT 10
PCT-US95-07820-8
; Sequence 8, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szareszen, Jędrzej B.
; APPLICANT: Szczygiowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; PCT-US95-07820-8

Query Match          9.1%; Score 220; DB 4; Length 63;
Best Local Similarity 67.3%; Pred. No. 5.1e-16;
Matches 37; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 332 IVSWCSQLEVLHAHPALGCFVTHCGWNSAVESLSCGVPVAVPQWFDQTTNAKLIE 386
Db 9 VVPWCPQLDVLHAHPAVGCFVTHCGWNSTLEALSFGVPMVAMALWTDQPTNARNE 63

RESULT 11
PCT-US92-00282-7
; Sequence 7, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DAREY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-00282-7

Query Match          6.9%; Score 166.5; DB 4; Length 529;
Best Local Similarity 22.1%; Pred. No. 6.5e-09;
Matches 77; Conservative 66; Mismatches 134; Indels 71; Gaps 16;

QY 121 RESQVPSALLWVEPA---TVLCIY-----YFYENGVADEID-----A 154
Db 139 RENQFDA--LFTDPAMPCGVILAEYLLKLPISYILFRGPPCSLEHIGQSPSPVSVYVFRFTK 196
QY 155 GSDEIQLP-RLPPLPQRSLSPTFLPETPERFLMMKEKLETLGDGEAKVLYNTFDALP 213
Db 197 FSDHMTFPQRLANFIENLYNLYHCLYSKYEILASDLL-----KRDV---SLPALHQ 246

```



```

Qy 214 DALTADRYELI--CIGPLIPSAFLDGGDPSETSYGGDLFEKSEENNCEVMDTKPKSSV 271
      ::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 247 NSWLLE-RYDFVFEYFPRVPMNMIFIGG--TNCKKGNLSQEF-----AYVNASGEHGI 298
      :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 272 VYVSFGS-VLRFPKAQMEIEIGKLLACGRPFLWMIREQKNDGCEBEELSCIGELKKMG 330
      |||||::||::||::||::||::||::||::||::||::||::||::||::||
Db 299 VVFSGLSMYSEIPEKKAMEIAEALGRIPQTLLWRYTGTGPSN-----LAKNT 345
      |||||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 331 KIVSWCSQLEVLHAPALGCCFYTHCGWNSAVESLSCGVPVAVPQWFDTTNAKLIEDAWG 390
      :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 346 ILVKVLPQNDLLGHFKARAFITHSGSHGIVYEGICNGVPMVMPFPGQMDNAKMETR-G 404
      :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 391 TGVVRMEHGGGVGDSEICERCEVYMDGGSEKSLYRENAIKWKTILARE 438
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 405 AGVTNLNLE--MTADDLENALKTVIN---NKSYPENIMRLSSLHKD 445
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 12
PCT-US92-00282-6
: Sequence 6, Application PC/TUS9200282
: GENERAL INFORMATION:
: APPLICANT: OWENS, IDA S.
: APPLICANT: RITTER, JOSEPH K.
: TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
: TITLE OF INVENTION: THEREIN.
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DARBY & CUSHMAN
: STREET: 1615 L STREET, N.W.
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20036-5601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/00282
: FILING DATE: 19920110
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: SCOTT, WATSON T.
: REGISTRATION NUMBER: 26581
: REFERENCE/DOCKET NUMBER: 91532-PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-861-3000
: TELEFAX: 202-822-0944
: TELEX: 6714627 CUSH
: INFORMATION FOR SEQ ID NO: 6:
: LENGTH: 531 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US92-00282-6

Query Match 6.6%; Score 160.5; DB 4; Length 531;
Best Local Similarity 23.5%; Pred. No. 2.8e-08;
Matches 52; Conservative 40; Mismatches 80; Indels 49; Gaps

Qy 229 PLIPSAFLDGGDPSETSYGGDLFEKSENNCV-----EWLDTKPKSSVYVSFGS 378
      ::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 265 PIMPNMVFIGG-----INCVIKPLKSQEFAYVNASGEHGI VVFSGLS 357
      :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 279 -VLRFPKAQMEIEIGKLLACGRPFLWMIREQKNDGCEBEELSCIGELKKMGKIVSWCS 337
      :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 308 MVSEIPEKKAMEIAEALGRIPQTLLWRYTGTGPSN-----LAKNTILVKWLP 354
      :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 338 QLEVLHAPALGCCFYTHCGWNSAVESLSCGVPVAVPQWFDTTNAKLIEDAWGTGVVRM 397

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2000, 21:35:38 ; Search time 9137.45 Seconds
(without alignments)
294.342 Million cell updates/sec

Title: US-09-147-955-1
Perfect score: 1507
Sequence: 1 gaaatttcacaaaaatgg.....ttataatttttagtaacaaaa 1507

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: em_fun.*
- 17: em_hum1.*
- 18: em_hum2.*
- 19: em_in.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_sy.*
- 29: em_un.*
- 30: em_vi.*
- 31: gb_htg1.*
- 32: gb_htg2.*
- 33: gb_in1.*
- 34: gb_in2.*
- 35: em_ba1.*
- 36: em_ba2.*
- 37: em_hum3.*
- 38: em_hum4.*
- 39: gb_pr4.*
- 40: gb_htg3.*
- 41: gb_htg4.*
- 42: gb_htg5.*
- 43: gb_htg6.*

- 44: gb_htg7.*
- 45: em_htg1.*
- 46: em_htg2.*
- 47: em_htg3.*
- 48: em_hum5.*
- 49: gb_pl3.*
- 50: gb_pr5.*
- 51: gb_htg8.*
- 52: gb_htg9.*
- 53: gb_htg10.*
- 54: gb_htg11.*
- 55: gb_htg12.*
- 56: gb_htg13.*
- 57: gb_htg14.*
- 58: gb_in3.*
- 59: gb_htg15.*
- 60: gb_htg16.*
- 61: gb_htg17.*
- 62: em_htg4.*
- 63: em_htg5.*
- 64: em_htg6.*
- 65: em_htg7.*
- 66: em_hum6.*
- 67: gb_htg18.*
- 68: gb_htg19.*
- 69: gb_htg20.*
- 70: gb_htg21.*
- 71: gb_htg22.*
- 72: gb_htg23.*
- 73: gb_htg24.*
- 74: gb_htg25.*
- 75: gb_htg26.*
- 76: gb_htg27.*
- 77: gb_htg28.*
- 78: gb_htg29.*
- 79: gb_htg30.*
- 80: gb_htg31.*
- 81: gb_vi1.*
- 82: gb_vi2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1507	100.0	1507	7	AB013596
2	1329.4	88.2	1458	7	AB013597
3	700.2	46.5	2043	7	AB013598
4	455.4	30.2	1864	7	AB027455
c 5	300	19.9	198944	8	ATCHRIV38
c 6	300	19.9	200576	7	ATFCA0
7	283	18.8	1594	7	AB000623
c 8	222.8	14.8	84203	49	AC005106
c 9	222.8	14.8	103223	8	AC007153
c 10	184.4	12.2	195068	8	ATCHRIV42
c 11	184.4	12.2	197419	8	ATCHRIV41
c 12	184.4	12.2	205065	7	ATFCA4
c 13	181.2	12.0	1589	8	ATU81293
14	160.8	10.7	99188	49	AC006533
15	159.6	10.6	1513	5	A62523
16	159.6	10.6	1513	5	A72396
17	149.2	9.9	1588	5	A62526
18	148.4	9.8	1459	5	A62525
19	145.2	9.6	1969	5	A62529
20	137.2	9.1	122358	8	AC002396
21	132	8.8	80818	7	AB019232
22	131.8	8.7	1416	7	MZEIAGLU
23	131.8	8.7	94786	49	AC002333
24	127.4	8.5	1440	49	AF190634
					AB013596 Perilla f
					AB013597 Perilla f
					AB013598 Verbena h
					AB027455 Petunia x
					AL161538 Arabidops
					297335 Arabidops
					AB000623 Nicotiana
					AC005106 Genomic s
					AC007153 Arabidops
					AL161542 Arabidops
					297339 Arabidops
					U81293 Arabidops
					AC006533 Arabidops
					A62523 Sequence 28
					A72396 Sequence 28
					A62526 Sequence 31
					A62525 Sequence 30
					A62529 Sequence 34
					AC002396 Arabidops
					AB019232 Arabidops
					L34847 Zea mays IA
					AC002333 Arabidops
					AF190634 Nicotiana

Db 841 GTTGGAGCGTTTGGAGTTTCCAAAGGCACAAATGGAAGAGATTGGGAAGGGCTATT 900
QY 901 agcctcggaagccgcttttctatgatgatacagacaagaatacagcagcggaaga 960
Db 901 ACCCTGCGAAGCGCGTTTATGGATGATACGAGAACAGAAATCAGCAGCGGGAAGA 960
QY 961 agaaagaagaagttgagttgcatctggggaatgaaaaaatgggaaaatgatttcgtg 1020
Db 961 AGAAGNAGNAGACTTCAGTTGCATTGGGGAATTGMAAAAATGGGAAAATAGTTTCGTG 1020
QY 1021 gtctgcagttggaggttcttggcgacacctgggttgggattgttctgacgaattgtgg 1080
Db 1021 GTGCTCGCAGTTGGAGGTTCTGGCGCACCCCTGCCGTTGGGATGTTGCTGACGCAITGTGG 1080
QY 1081 gtggaactcggcgttggagagcttgagttcgagggttcagggttggtcggtccagctg 1140
Db 1081 GTGGAACTCGCGCTGGAGAGCTTGAGTTGGCGGTTCCGGTGTGGCGGTGCCGAGTG 1140
QY 1141 gttgatcagacgacgaatgcgaagctgattgaggatgctgggggacaggggtgagagt 1200
Db 1141 GTTGTATCAGACGACGAATGCGAAGCTGATTGAGGATGCGTGGGGACAGGGGTGAGAGT 1200
QY 1201 ggaatgaataaaggggtgggttgatgatctgacatagacagatggtgtaaatggt 1260
Db 1201 GAGATGAATGAAGGGGTGGGTTGATGATCTGAGATAGAGAGTGTGTGGAGATGT 1260
QY 1261 gatgatgggggtgagaagagcaactagtagagaaaaatgccataaaatggaagacttt 1320
Db 1261 GATGGATGGGGTGAGAAGACAACTAGTGAGAGAAATAGCCATAAAATGGAAGACTTT 1320
QY 1321 ggcagagaagccatgggagaggtgagatcttcaactcaagaatcacaagccttcttca 1380
Db 1321 GCGCAGAGAACCATGGGAGAGGATGGATCTCACTCAAGANTCTCAAGCCCTTCTTCA 1380
QY 1381 tcaagtcacagctgcttaatacacaataatggcttccacttcaacttaactcaacacagc 1440
Db 1381 TCAAGTGCACGTGCTTAATACACAAAATGGCTTTCCACTTTTAATCTCAACACAGC 1440
QY 1441 gtccaaaataatcccttccacttcttcttcttcttcttcttcttcttcttcttctt 1500
Db 1441 GTCAAAATAAATATCCCTTCCACTTCTTCTATTCTTCACTATCACATTTATATTAGT 1500
QY 1501 acaaaa 1507
Db 1501 AAAAAA 1507
RESULT 2
AB013597
LOCUS AB013597 1458 bp mRNA PLN 20-MAR-1999
DEFINITION Perilla frutescens PF3R6 mRNA for UDP-glucose:anthocyanin
5-O-glucosyltransferase homologue, complete cds.
ACCESSION AB013597
VERSION AB013597.1 GI:4115560
KEYWORDS PF3R6; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.
SOURCE Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to
mRNA.
ORGANISM Perilla frutescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Gentiananae; Lamiales; Lamiaceae; Perilla.
REFERENCE 1 (sites)
AUTHORS Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,
Kusumi,T. and Saito,K.
TITLE Molecular cloning and biochemical characterization of a novel
anthocyanin 5-O-glucosyltransferase by mRNA differential display
for plant forms regarding anthocyanin
J. Biol. Chem. 274 (11), 7405-7411 (1999)
MEDLINE 99167509
REFERENCE 2 (bases 1 to 1458)
AUTHORS Yamazaki,M., Saito,K. and Gong,Z.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami

Yamazaki, Pac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
Biotech.; Yayoi-cho i-33, Inage-ku, Chiba, Chiba 263-8522, Japan
(E-mail:mami@p.chiba-u.ac.jp. Tel:+81-43-290-2305,
Fax:+81-43-290-2305)
Location/Qualifiers
1. .1458
/organism="Perilla frutescens"
/cultivar="Shikun"
/variety="crispa"
/db_xref="taxon:48386"
/tissue_type="leaf"
29. .1360
/gene="PF3R6"
29. .1360
/gene="PF3R6"
/codon_start=1
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase
homologue"
/protein_id="BAA36422.1"
/db_xref="GI:4115561"
/translation="MVRRLVLLATFPAQGHINPALQFAKRLLKAGTDVTFSTVYAWR
RWANTASAAAGNPGLDFVAFSDGDDGLKPGDGKRYMSEMKARGSEALRNLNND
DVTFVYSHLEFAWAAYARLSHVPTALLJWEPATVLCIYFVNGYADEIDAGSNEIQ
LPRPLSEORSLPTFLPATPERFLMKKLETLDEKAKVLYNTDALEPDLALTA
IDRELIGPLIPSAFLDGDSETSYGGDLFEKSENNCVENLNSPKSVYVTF
GSURFPKAAHEEIGKGLLACGRPLFMIREQKNDDBEEEBEELSCIGELKRMKI
VSWCSQLEVLAHALPAGCFVTHCGNSAVESLSCGIPVAVPQMFQDTTNKLIEDAWG
TGVRYRMNEGSGVDGCEIERCEVMVDGGDKTKLVRENAIKWTLARQMG"
1458
polyA_site
BASE COUNT 353 a 348 c 438 g 319 t
ORIGIN
Query Match 88.2%; Score 1329.4; DB 7; Length 1458;
Best Local Similarity 95.4%; Pred. No. 5.4e-280;
Matches 1403; Conservative 0; Mismatches 36; Indels 32; Gaps 2;
QY 2 aaatttcacaaaaatggtccgcgcgctgctgctagcaacgttctctcgcaagc 61
Db 14 AAAATTTCCAAAAAATGGTCCGCCGCGTGTGCTGTAGCAAGTTTCGGGCGAAGC 73
QY 62 cacataaatcccgccctccaattccgcaagagactcctataaagccggcaactgagtcaca 121
Db 74 CACATAAATCCCGCCCTCCAAATTCGCCAAGAGACTCTCTAAAAGCCGCGCTACGTCACG 133
QY 122 tttttcacagcggtttatgcatggcgccgcatggccacaacagacctccgcgctgcccga 181
Db 134 TTTTTCACGAGCGTTTATGTCATGGCGCGCATGGCCAAACACAGAGCTCGCGCGTCCGGA 193
QY 182 aaccaccgggctcgacttcgtggttctccgacggctacgacgagcggtcgagcc 241
Db 194 AACCCACCGGGCTCGACTTCGTGGCGTTCTCCGACGGCTACGACGAGCGGTGAAGCCC 253
QY 242 tcgscgagcgaagcgcgtacatgccagagatgaagcccgccgctcgagggcttaaga 301
Db 254 GCGCGGACGGGAGCGCTACATGTCGAGATGAAGCCCGCGCTCGGAGGCTTTAAGA 313
QY 302 aaactctctcaacaacacagcgtcacgttcgttcgttcactccaccctcttgcattg 361
Db 314 AACCTCTCTCAACAACGACGACGTCATCTTTCGTCGCTACTCCACCCTCTTTGTCATGG 373
QY 362 gcggcgaggtgagcgtgagtcgccaggtcccgagcaccctctctggtcgagccgccc 421
Db 374 GCGCGGAGGTGGCGGTTTGTCCACGCTCCGACCGCCCTCTCTGGGTCGAGCCGCC 433
QY 422 accgtgctgcatatattacttctacttcaacggctacgacgacgagatcgagccggt 481
Db 434 ACCGTGCTGTCATATACCACTCTTACTTCAACGGCTACGACGAGATCGACGCGGT 493
QY 482 tcgcagaaattcagctccctcggttcacccctcgagcagcagcagcttccgaccttt 541
Db 494 TCCAATGAATTCAGTCCCTCGGCTTCATCCCTGGAGACGCGCATCTTCCGACGTTT 553


```
QY 301 -----aaacctctctcaacaacacgaagcagcgtctcgtctac 343
Db 302 GATACTCTTGCAGCAATAATGTCGATCAAAAAGCAGCAAAATCACGTTCTGTTGATAC 361
QY 344 tccacacctcttgcagcggcgaggtggcgctgagtcacaggtcccgagcgccctt 403
Db 362 TCCCACTCTTTCATGCGCGCAAGTGGCGGTGAGTCCATCTCCGAGCGCGTA 421
QY 404 ctctgggtcgagccgcacacgctgctgcatatattacttacttaacagcgtacgca 463
Db 422 CTCTGGATGACCCAGCTACCGGTCTTGGATATATTTTACTTTTATTTCAACGGCTATAGC 481
QY 464 gacgagatcgacgcgggttcgcagcaaatca---gctccctcggtccacccctcgag 520
Db 482 GACGAAATCGATGCGGGTTCGATGCTATTCACTTCCCGGAGGACTCCACGTCGTGCC 541
QY 521 cagcgcaatttcgacaccttctcctccggagacacccggagagatcccggttgatg 580
Db 542 CAGCGTGATTTACCGCTCTTCTTCTTCCCTCCACGATGAGAGATTCGTTCTACTGATG 601
QY 581 aaggagaagctggaaacttttagacggtgaagagaagcgaaagtgttggtgaacacgttt 640
Db 602 AAGGAGAAATTGGAACCTTTAGAAGGTGAAGAAACCTAAGCTCTTGTGTAACAGCTTT 661
QY 641 gatcggttgagcccgatgcaactcaacgctatgtagtagtagtgatgagtcggatcg 700
Db 662 GATGCGTTGGAGCTGATGCGCTCAAGGCCATTGATAAGTACGAGATGATTCATTCGG 721
QY 701 ccgttgattccctcgaccttcttgacgcgggagatccctccgaaacgtcttaacggcgc 760
Db 722 CGGTTGATTCCTTCGCACTTCTTGGACGGTAAAGATCTTTCGACAGGCTTTCGCGGA 781
QY 761 gatcttttcgaaaaa---tcgagagagaataactgctggagtggttgagacgaagccg 817
Db 782 GATTGTTTCGAGAAAGGTGCAATGACGAGATTCGCTCGAATGTTGAGCACGAATCCT 841
QY 818 aaatctcgggtggtatgctcgtctggagcgttttgaggtttccaaaggcacaatg 877
Db 842 CGATCTCGGTGTTACGTTTCGTCGGAAGCTTCGTTAATACGACGAAGTCGCAATG 901
QY 878 gaagagattggaaaggctattagcctgcggaagccgctttttagatgatacgagaa 937
Db 902 GAAGATAGCAAGAGGCTCTTATGATGTGGAGGCCGTTTTTGGGTGTGTAAGACTA 961
QY 938 cagaagaatgacgcgcggaagaagaagaagagttgagttgcatggggaattgaaa 997
Db 962 AA-----CGAAGGAGAGAGGTATTGATAAGTTGCATGCGAGGTTGAAA 1006
QY 998 aaaaaggagaaatagttctgctcgaagtggaggttctgcgcacacctcggtg 1057
Db 1007 CGAGTGGGGAATTTGATCTGTGTTCTCAATTTGAAGTCTCGACGATCCCTCGTGTG 1066
QY 1058 ggaatgttcgtaacgcatgttggtggaactcggctgtgagagcttgagttgcggggtt 1117
Db 1067 GGATGTTTCGTGACACACTCGCGGTGGAATTCGACTTAGAGAGTATATCTTTTCGGGTT 1126
QY 1118 ccggttggtcgcggtgacgagtttgatcagacacgaatacgaactgattgagat 1177
Db 1127 CCGATGGTGGCTTTTCCGCACTGTTTCGATCAAGGAGGAGNAATCGAAGCTGATGAGGAT 1186
QY 1178 gcgtgggggacaggggtgagagtgagaatgaatgaaggggggt---ggggttgatgatct 1234
Db 1187 GTGTGGAGGACGGGTGTGAGAGTGAGAGCTAATGAGGAGGGGTAGCGCTGCTGATGATG 1246
QY 1235 gagatagagaggtggtgagatggttatgagtggggggtgagaagaagcaaacagttaga 1294
Db 1247 GAAATTAGGAGATGATTGAGGAGGTTATGGATGGGGGAGAAAAGAGTAGGAACATTAGA 1306
QY 1295 gaaaatgccataaaatgaaagactttggccagagagaagccatggagagagatgatcttca 1354
Db 1307 GAGATGCTGGCAAGTGGAGGATTTGGCAAGAAAGCATATGGAGGAACATGATGATCTTCA 1366
```

```
QY 1355 ctcaagaatctcaacgcctttcttctcaatcaagttgcagctgcttaataacaaatgctt 1414
Db 1367 GTTAAACACTCAAGGCTTCTTCTTGTATGAGGTTGTAGGTATCTAAAGACGTAATAGGT 1426
QY 1415 tcc 1417
Db 1427 CCC 1429

RESULT 4
LOCUS AB027455 1864 bp mRNA PLN 07-JAN-2000
DEFINITION Petunia x hybrida PH1 mRNA for anthocyanin 5-O-glucosyltransferase, complete cds.
ACCESSION AB027455
VERSION AB027455.1 GI:6683051
KEYWORDS anthocyanin 5-O-glucosyltransferase.
SOURCE Petunia x hybrida (cultivar:Surfinia) corolla cDNA to mRNA.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE 1 (bases 1 to 1864)
AUTHORS Yamazaki,M. and Saito,K.
TITLE Anthocyanin 5-O-glucosyltransferase
JOURNAL Published Only in Database (2000) in press
REFERENCE 2 (bases 1 to 1864)
AUTHORS Yamazaki,M. and Saito,K.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences, Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan (E-mail:mami@p.chiba-u.ac.jp, Tel:81-43-290-2905, Fax:81-43-290-2905)
FEATURES
     source
       1..1864
         /organism="Petunia x hybrida"
         /cultivar="Surfinia"
         /db_xref="taxon:4102"
         /tissue_type="corolla"
         100..1506
           /gene="PH1"
           100..1506
             /gene="PH1"
             /codon_start=1
             /product="anthocyanin 5-O-glucosyltransferase"
             /protein_id="BAA89009.1"
             /db_xref="GI:6683052"
             /translation="MVQPHVILTFPAQGHINPALQFAKLVKMKIEVTFSTSIYAOS
             RMDKSLINAPKGLNFIFFSDGDEGDFDSKDPVFMQLKCGSETVKKILLTCSEN
             GQPTCLLLSIFLPWAAEVAEHVIFSAIWSOPATILDIITIFNPHGTEKAWANESND
             PNWSIPLGPLELTDRLDPLSPFLPYGAKSLRVALPPFKELIDTLDATTPKILVNTF
             DELEPEALNAIEGYKFGIGLIPSPALGNDPLDASFGDGLFQNSNDYMLNLSKPE
             SSVYIISFGLMNPISQMBEISKGLIDIGRPLWVFKENKGEKEENKKGICIELE
             KIGIVPWCQLEVLKHPISGCFVSHGWSALESLACGPVAVPQWTDQNTAKQV
             EDVWKSQVVRVINEDVVESEIKRGTIELVMDGGEKELRNKAKKRELAREAVKEG
             GSSHNKLFATDIDVAKCF"
BASE COUNT 603 a 318 c 369 g 574 t
ORIGIN
```

```
Query Match 30.2%; Score 455.4; DB 7; Length 1864;
Best Local Similarity 61.6%; Pred. No. 2.6e-89;
Matches 871; Conservative 0; Mismatches 491; Indels 51; Gaps 7;
```

```
QY 13 aaaaatggtcccgcccgctgctgctgagcaacgttttctcgcaagggcacaataatcc 72
Db 96 AAATATGTTGTCAGCTCATGTCATCTTTAAACAACATTTCCAGCACAGGCCATATTATCC 155
QY 73 cgccctcccatctcgcaagagactcctaaagccggcagctgaactgacatcattttcacgag 132
Db 156 AGCACTCAATTTGCCAAGAACTCTTGTCACATGGGCATAGAGTGACATTTTCTACAAG 215
```

QY 133 cgtttatgcatgagcgccgcatgagcgaacacagcctccgcgcgtccgagaaacccaccg 192
Db 216 CATTATGCCCCAAGCGTATGATGATAAATCCATTCCTTAATGC-----ACCAAAAGG 269
QY 193 cctcgactcgtgcttctccgacggtcagcagcagcggtcgaagccctgcgcgcagc 252
Db 270 ATTCAATTTTCATTCCTATTTCCGATGGCTTTGATGAAGGTTTGATCAATTCAAAAGACC 329
QY 253 gaacgctacatcgcagatgaaagccgcgcgcgtccgagcccttaagaacacctctct 312
Db 330 TGTATTTACATGTCACAACTTCGTAATGTGGAAGTGAACACTGTCAAAAAATAATTC 389
QY 313 c-----aacaacacacagctcagcttcgctcactcctccacctcttgcag 360
Db 390 CACTTGCTCTGAAAATGGACAGCCTATACTTGCCTACTTACCTCCATTTTCTCTCTG 449
QY 361 ggcgcgagaggtgagcgtgagtcaccaggtccgcgcgcgcctctctctggtcgagccgc 420
Db 450 GGCAGCAGAGGTACACGTAAGTTCACATCCCTTCTGCTCTCTTTGGAGTCAACACGC 509
QY 421 caccgctgctgcatatattactctcactcaccgctcagcagcagcagcagcgcgcg 480
Db 510 AACAAATTGGACATATATTACTTCAACTTTCATGGATGATAAAGCTATGGCTAATGA 569
QY 481 ttcgcagcaa-----attcagctccctcgcgtccctccacctcgtgagcgcgcag 528
Db 570 ATCCAATGATCCAAATTTGGTCCATTCACATCCCGGGCTCCACTACTTGGAAATCCAGA 629
QY 529 tctccgacacctctctgcgc-----ggagacacccgcgagagattccgggttgatg 580
Db 630 TCTTCTTCTATTTTACTTCTCTTATGGTCAAAAGGGAGTCTTCAGTTCACATTCAC 689
QY 581 ----aagagaagctgaaacattcagctgtaagagaagcgaaagtgttggtgaacac 636
Db 690 ATTCAAGAATTGATAGACACATATAGTCTGAACCACTCTAGATATCTTGTGAATAC 749
QY 637 gttgatgcgttggagccgatgacactcgcgctattgataggtgatgagttgatcgag 696
Db 750 ATTTGATGAATTAGAGCCTGAGCAGCTCAATGCAATTAAGGTTTATAAGTTTATGGA 809
QY 697 cggcccttgcctccgcctctctgacgcgcgcgcgcgcctccgcgaacgtcttcag 756
Db 810 TGGACCGTGTATCTCTCTCTTCTTGGGTGGAATGACCTTTAGATGCTTCATTTGG 869
QY 757 cggcgatcttttcgaaaaatcggagagagataactcgcgtggagtggttgggacgaagcc 816
Db 870 TGGTGAATCTTTTCAAAATTCAAATGACTATA-----TGGAAATGGTTTAAACTCAAAAGCC 923
QY 817 gaaatctcgtggtgtagtgcgttctggagcgcttttgaggtttccaaaggccacaa 876
Db 924 AAATTCATCAGTGTATTATATATCTTTGGAGTCTAATGAATCCATCTATTAGCCAAAT 983
QY 877 ggaagagattggaaaggcctattagcgcgcgcgcgcgcgcctttttatgatgatacaga 936
Db 984 GGAGGAGATATCAAAAGGGTTGATGACATAGGAAGCGCGTCTTTTATGGGTGATAAAGA 1043
QY 937 acagaagaatgacgc 996
Db 1044 AAATGAAAAGGCCA-----AGAGAAGAGAAATAAAGAGCTGGTGTATTGAAGAAATGGA 1100
QY 997 aaaaatgggaaaaatagttctcgttcgtcgcagttgaggttctgcgcgcgcgcgcgcgc 1056
Db 1101 AAAAATAGAAAAATAGTTCATCGGTGTTCACAACTTGAAGTCTTAAACATCATCTCTTT 1160
QY 1057 gggatgttcgtgacgcagtggtggagactcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1116
Db 1161 AGGATGTTTGTCTCATCTGCTGGATGGAATTCAGCCCTTAGAGATTTAGCTTTGTGAGT 1220
QY 1117 tccggtggtgagc 1176
Db 1221 GCCAGTTGTGGCAATTTCTCAATGGACAGATCAAAATGCAAAATGCAAAACAAAGTTGA 1280
QY 1177 tgcgtgaggggacaggggtgagagtgagaatgaatgaaggggggtgaggtgagctctga 1236

Db 1281 TGTGTGGAAGATGGAGTAAAGTGAATAAATGAAGATGTTGTTGAAATGAGGA 1340
QY 1237 gatagagagtggtgagatggtgagtggtggggtgagaagagcaaacctcagtgagaga 1296
Db 1341 AATCAAAAGGTGATTGAATTTGTAATGATGATGAGAGAGAAAGGGGAATTTGAGAAA 1400
QY 1297 aatgcgcataaatggaagactttgccaagagaagccatgggagagatgagcttcaact 1356
Db 1401 GAATGCTAAAGAAATGGAAGAATGGCTAGAGAAGCTGTGAAGGAAGGTGATCTTACA 1460
QY 1357 caagaatctcaacgcctttcttcacatcaagtgc 1389
Db 1461 CAAGAATTTAAAGCCTTTTATTGATGATGTTGC 1493
RESULT 5
ATCHRIV38/c
LOCUS ATCHRIV38 198944 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 38.
ACCESSION AL161538
VERSION AL161538.2 GI:7268110
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 198944)
AUTHORS EU Arabidopsis sequencing, project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV37 at the 5' end and an
overlap with ATCHRIV39 at the 3' end.
FEATURES
Location/Qualifiers
source 1. .198944
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
4845..5598
/gene="AT4g14060"
join(4845..5031,5330..5598)
/gene="AT4g14060"
/note="strong similarity to major latex protein type 1,
Arabidopsis thaliana; contains EST gb:A1994471.1, T43693,
AI099764, H36765, T43905, AI998305.1, R29829, T22625,
R90142, T41926, T22170, W43041, R65053, AA712594,
AA712597."
/codon_start=1
/product="major latex protein like"
/protein_id="CA878448.1"
/db_xref="GI:7268111"
/translation="MATSGTYVTEVPLKSAEKHYKWRNENHLFPDAIGHHIQGVTV
HDGEWSHGALKIWNVTLDGKPEKPEREDDENMAVTFVLEGRHMEOLKVDYDIL
QFLQKSPDDIVCKITMTWEKRADDSPEPSNYMKLVKSLAAMDHAHLKA"
4845..5031
/gene="AT4g14060"
/number=1
5032..5329
/gene="AT4g14060"
/number=1
5330..5598
/gene="AT4g14060"
exon
intron
exon


```
exon      /number=2
gene      complement(9562. .9654)
CDS      /number=1
          9562. .15425
          /gene="AT4g14070"
          complement(join(9562. .9654,9750. .9807,9883. .9998,
10097. .10176,10276. .10435,10594. .10731,11065. .11145,
11247. .11300,11383. .11541,11653. .11758,11882. .11949,
12156. .12305,12439. .12554,12626. .12834,13288. .13360,
13988. .14072,14674. .14859,15291. .15425))
/notes="strong similarity to AMP-binding protein - Brassica
napus; Contains putative AMP-binding domain signature
AA255-266; contains EST gb:AI999282.1"
/codon_start=1
/product="AMP-binding protein"
/protein_id="CAB78449.1"
/db_xref="GI:7268112"
/translation="MASTSGSLPSTLLSGVSRPDPGFRLLSGHESVRIPSRFP
REKVKSSPFLSSFSGDALRSSEKAWKADPDWRSSAEKYGDRVALVDYHDPPLK
LTYQLQEQEILDFEAGLRVGVKADSKIALFADNSRWLSDQGANVVRGSRSSVEE
LQIYRSESVAIVDPEFNRIARFSTSKASRLFLILLAGEKSSLVTCQMQLPVYS
YAEIINGQESRAKLSASNDTSRYRNOFTDSDDTAAIMTSGTGNPKGVLTHRNLL
HOIKHLSKYPAAQDKPFLSMLPSHAYRASEYFIETCGVEOMYTSRYLKDCLKRY
QNTIVSPVLAVETLYSGIQKQISASSACKKFLATLLKVSMAVMEMKRIYEGCLJK
EQKPMXIVAPDLNARVIAALLPFLHMLAKKLIYKKIHSISIGSKAGISGSLPI
HYDFFAIGVILQNGTETSPVVCARTLSCNVGSAGHPMHGTETKIVDPETNNV
LPPGSGKILKRVGVQVMKGYKNPSTTKVLNESGWFNGTGTNAPHHSKGRHCG
GVIVLEGRADTVLSTGENVEPLEIEEAMRSRVIEQIVIGQDRRLGALIIIPKE
EAQVDPETSKETLKSILVQELRKVTSEGSFQVGPVLIVDDPFTIDNGLMPTTKIRR
DMVAVKKEEIDQLYS"
          complement(9655. .9749)
          /number=1
          complement(9750. .9807)
          /number=2
          complement(9808. .9882)
          /number=2
          complement(9883. .9998)
          /number=3
          complement(9999. .10096)
          /number=3
          complement(10097. .10176)
          /number=4
          complement(10177. .10275)
          /number=4
          complement(10276. .10435)
          /number=5
          complement(10436. .10593)
          /number=5
          complement(10594. .10731)
          /number=6
          complement(10732. .11064)
          /number=6
          complement(11065. .11145)
          /number=7
          complement(11146. .11246)
          /number=7
          complement(11247. .11300)
          /number=8
          complement(11301. .11382)
          /number=8
          complement(11383. .11541)
          /number=9
          complement(11542. .11652)
          /number=9
          complement(11653. .11758)
          /number=10
          complement(11759. .11881)
          /number=10
          complement(11882. .11949)
          /number=11
          complement(11950. .12155)
          /number=11
          complement(12156. .12305)
          /number=11
          /number=2
          complement(12306. .12438)
          /number=12
          complement(12439. .12554)
          /number=13
          complement(12555. .12625)
          /number=13
          complement(12626. .12834)
          /number=14
          complement(12835. .13287)
          /number=14
          complement(13288. .13360)
          /number=15
          complement(13361. .13957)
          /number=15
          complement(13958. .14072)
          /number=16
          complement(14073. .14673)
          /number=16
          complement(14674. .14859)
          /number=17
          complement(14860. .15290)
          /number=17
          complement(15291. .15425)
          /number=18
          complement(16137. .16174)
          /number=1
          16137. .17732
          /gene="AT4g14080"
          complement(join(16137. .16174,16252. .17604,17687. .17732))
          /note="the cDNA A26668 is cognate to this gene; gene A6
          for anther-specific protein - beta-1, 3-glucanase homolog;
          Contains Glycosyl hydrolases family 17 signature
          AA274-287; Prokaryotic membrane lipoprotein lipid
          attachment site AA8-18; contains EST gb:AW004339.1,
          Z18400"
          /codon_start=1
          /product="A6 anther-specific protein"
          /protein_id="CAB78450.1"
          /db_xref="GI:7268113"
          /translation="MSLLAFETILVFSSCCSATRFQGHRYMORKTMDLASKIGI
          NYGRGNLPSPYGSINFIKSRAGHKWLDADPESITLLSOTNLIVITVPHNHOITA
          LSSNOTIADEWVRTNIIYPQTRFVLVGNELISYNSGVSNLVNLPANRIVNSLR
          LRGHINIKVGTPLAMDSLRSPNSFPNGTFRREITGPMPLPKFLNGNSYFVLNVRP
          YFRWSRPNMTSIDETALFQGHSTYDPTQGLVYRNLLDQMLDSVLFAMTKLGYPHMLR
          AISEGTWPNFGDIDETCANILNAATYNNLKKMSASPPITGTPSRPGLPIPTFVPSLF
          NENOKSGSGTORHWGILHPDGSPIYDVTGQPLGPNPLPKPTNNVYKGVQWCVF
          VEGANETELETLRMACAQSNTTCAALAPGRCYEPVSIYWHASYNALSYWAQFRNQS
          IOCFNGLIAHETTTNPGNDRCKSPSVTL"
          complement(16175. .16251)
          /number=1
          complement(16252. .17604)
          /number=2
          complement(17605. .17686)
          /number=2
          complement(17687. .17732)
          /number=3
          18189. .18494
          /gene="AT4g14090"
          /number=1
          18189. .25761
          /gene="AT4g14090"
          join(18189. .18494,18694. .18787,18938. .19018,19571. .19724,
19772. .19833,21489. .21524,22014. .22792,22844. .22948,
23060. .23200,23280. .23566,24375. .24631,24686. .25140,
25192. .25311,25445. .25614,25665. .25761)
          /gene="AT4g14090"
          /note="similarity to proton-coupled peptide transporter
          PEPT, Rattus norvegicus; Contains Prokaryotic membrane
          lipoprotein lipid attachment site AA535-545; Prokaryotic
          membrane lipoprotein lipid attachment site AA955-965;
          contains EST gb:AI998193.1"
```

Query Match	19.9%;	Score 300;	DB 8;	Length 198944;
Best Local Similarity	55.7%;	Pred. No. 2.5e-530;		
Matches 768;	Conservative 0;	Mismatches 55;	Indels 80;	Gaps

Qy	36	tgtatgaacggttctctgcgcaaggccacataaataccgcctcccaattcgcacaagac	95
Db	21201	TGCTTGTAACTTCCACGCGCAAGGTCACTCAACCCGCGCTCAACTAGCCACCGCC	21142
Qy	96	tctataaagccggcgactgacgtcacatttttcaagagcgtttatgatggcgccgcatg	155
Db	21141	TCATCCACACCGGTGCACCGTCACTACTCCACCGAGTCTTGCTCAACCGACGTATGG	21082
Qy	156	ccaacacagcctcccgctgcgcggaacccaccggcctcgcactcgtggcgattccgc	215
Db	21081	GCAGCGCACTTCCA-----CAAAAGGTCTATCTCTCGCTTGTTGTTCAACG	21037
Qy	216	acggctcagaacagcggtgaagccctgcggcgagcggaagcgcgtcacatgctccgagatga	275
Db	21036	ATGGATTGCAGACCGGTCTCAAGTCAATTTCGAAGACCAAGAAATCTACATGTCGAAC	20977
Qy	276	aagccgcggctccgagcgcttaagaacacctctctcaacaccacgcgcg-----	326
Db	20976	AACGATGTGTTCAAAACGCCCTGAGAGACATCATAAAGCCAATCTTTGACGCCAC	20917
Qy	327	-----tcagttctgcgtactccacacctcttgatggcgggcgcgaggtgg	374
Db	20916	AAACAGAGCTATCACCGGGTATCTACTCTGTTCTCGTCCGCTGGGTTTCACGGTAG	20857
Qy	375	cgcgtgagtcaccaggtcccgagcgccctctcttggtcgagcccgccacgctgctgtgca	434
Db	20856	CGCGTGAGTTTCACCTCCCACTACACTCTCTCGATTGAACCAAGCTACTGTACTAGCA	20797
Qy	435	tatatctcttacttcaacggctacgcagacgagatcgacgcggttccgcagaaattc	494
Db	20796	TCCTACTACTACTACTCAACACCTCTTCAACATCTCTTCGAGGTTGAACCGATTAAAT	20737
Qy	495	agtcctctcggcttccacccctggagcgcgagctcttcgacaccttctgctgcgcgaga	554
Db	20736	TACCGAAACTGGCACTGATCACACCGGTGACCTCCGCTGCTTTCTTCAACCTTCGAAG	20677
Qy	555	caccggagagattccggttgatgataagaggaagctcgaaactttagacggtgaagaaga	614
Db	20676	CATTACCGTCAGCTCTTTGACTCTAAGAGAACATATCGAAGGCTCTCGAACCGAATCAA	20617
Qy	615	aggcgaaagtgtggtgaacacggttgatgcggttgagcccgatgcactcaaggctatcg	674
Db	20616	ACCTTAAGATTCTGTATACACATTTCTGCTTTTGGAAACACGATGCTTTAACTCTGT	20557
Qy	675	ataggatgattgacgggacggcggttgattccctccgcctctcttgacgcgcgag	734
Db	20556	AGAACTCAAGATGATCCCAATCGACCGCTTGTTTCTTCTCGAGGGTAA-----	20503
Qy	735	atccctccgaagcttctacggcgcgactcttccgaaaaatcgggaggaatacactgcg	794
Db	20504	-----AACCGATCTTTTCAATCTTCGACGAGGAT---TACA	20470
Qy	795	tggagtgggttgacacgaagccgaaattctcggtgggtgatatgctgcttggga---	851
Db	20469	CGAAATGGTTAGACTCGAAGCTCGAGATCAGTGATTTACATTTCTTTAGGCACACAG	20410
Qy	852	ttttgaggtttccaaagccaaaatcgaaagagatgtgggaaagggtctattagctcgga	911
Db	20409	CCGATGATTTACAGAGAAACATGTGAAGCGCTTACTCAACCGCGTGTAGCTTACAAACA	20350
Qy	912	ggcgtttttatggatgatacagagacagaagaatgcacgcgcgaagaagaagaagaag	971
Db	20349	GACCGTTTTATGATCGTGAGGGAGAAAATCCAGAAG-----AGAAGAAAGAAGAT	20297
Qy	972	agttgagttgcattcggggaattgaaaaaaatcgggga--aaatagtttcgtggctgcgag	1030
Db	20296	CGGTTTCTTGAATTGATCAGAGAAAGTAGATGATCGAGGATTTGGTGGGATGTTCTC	20237
Qy	1031	ttggaggttctgcgcacccctgcgtgggattggttcgtgcagcattgtgggtggaactcg	1090

Db	20236	ACAGCTGTTTGGCGCATGTGCTGCGGATGTTTGAGACTCATTGGTGGAAATTCG	20177
Qy	1091	gctgtggagagcttgagtgcgggttcgggttccgggtggcggtgccgcagtggtttgatcag	1150
Db	20176	ACGTTGCAGAGTTTAGAGAGTGGTGTCCGCTGGTTCGCTTTCCGACGTTTGCTGATCAG	20117
Qy	1151	acgacgaatcgaagctgattgaggatcgctggggggaccagggtgagagtgagaatgaat	1210
Db	20116	TGTACAACGCCGAAGCTTGTGGAGGATACGTGGAGATTTGGAGTTGAAGTCAAGGTTCGG	20057
Qy	1211	gaagggggttgggttatgatactgagatcagagaggtgtgtggagatggtgatgagtg	1270
Db	20056	GAGCAAGCAGATGTGGATGGGAGGAGATTAGAAGCTGTTTCGACAAGTGATGAGTGGT	19997
Qy	1271	ggtgagagagcaactagtagagagaaaaatgccataaaatggaagaccttggccacagaaa	1330
Db	19996	GGAGAAGAGCGGAGGAGATGAGAGAGAAATGCAGAGATGGAGGCGGATGCTCTTGAT	19937
Qy	1331	gccatgggagagatgagatcttccaactcaagaatctcaacgccctttctcatcaetgtg	1388
Db	19936	CGCGCAGCGAAGGTGGACCGTCGGATTGTAATCTTTAAAGGTTTTGTGCAGGAGATG	19879
<hr/>			
RESULT	6		
ATFCAO/c			
LOCUS	ATFCAO	200576 bp DNA	PLN 28-JUN-1999
DEFINITION	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment		
NO.	0.		
ACCESSION	Z97335		
VERSION	Z97335.2 GI:5280985		
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eucalyptophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi		
AUTHORS	Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N., Kreis, M., Kavanagh, T., Entlan, K.D., Rieger, M., James, R., Puigdomenech, P., Hatzopoulos, P., Obermaier, B., Duesterhoft, A., Jones, J., Palme, K., Ansorge, W., Deisen, M., Bancroft, I., Mewes, H.W., Schueller, C. and Chalwatzis, N.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 200576)		
AUTHORS	EU Arabidopsis sequencing project.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuellemips.biochem.mpg.de, mayeremips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk		
COMMENT	On Jun 30, 1999 this sequence version replaced gi:2244747. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATFCAL at the 3' end. Location/Qualifiers 1..200576 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 1..63289 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="BAC TMU20K8" /1551..4428 /gene="dl33000w" Join(1753..3194,3261..4428) /gene="dl33000w"		
FEATURES	source		
	source		
	gene		
CDS			

```

/note="similarity to Cf-2.2, Solanum pimpinellifolium"
/codon_start=1
/product="disease resistance Cf-2 like protein"
/protein_id="CAB10171.1"
/db_xref="GI:2244748"
/db_xref="SPTREMBL:O23253"
/translators="MTIITWISLCLIPCLNSILVIADKLDLPDORDALLEPKNEFSIP
SPQSDMLLIQTTAKRNNTDCSNGGISCDBPKTGVVVELDLGNLGNSSIF
RLQHLGSELDNDLCTIPDSGNGFYRLVNLGLFGEIPFSLRSLSYLDLGL
SYNDLLEIDLSMGNLHLRLVLSLSCAFTKIGFSSIGNUTYLFDDLSWNYTGEL
PDSMGNLKSRLVNLHRCNFGKIPFSLGSLNLDLIDISKEFTSEGPSMSLNRL
PQJMLNLSSTNVLDSSNQKAMLPNSMSLSLEAFDISGNSGTTPSSLFML
FSLKLLDGLNFKTIGNITSSPNLQELIGENNIPTRPSILKVLGSLALSLS
PMDTGGIVDSIFSLKLSLSDLSGILNLSHSLPSSHMHLILSSCSQPKFL
ENQTSVLHLDISANOEGOVPELWRLPTLSFIASDNKFSGEIPRACVCEIGTIVLSNN
NFSGIPPCFISNKTLSILHLRNNLSLGVIPESLHGYLRSLDVGNSRLSQPKSL
INCSYLOFLAVENRINDPFSKLSLPLQLLVLSRNEFHGPIFSPGDSLSFSKLEF
FDSISNFGVLSRSDYFVGSVNSYSYDIDINTPGFTVVGDDSEFHKSVYLTIKGLN
MELVSGFETIKTIDVSGNLEGDIPESIGLIDELIVLNMNNAFTGHIPSLNLSN
LQSDLSQNLSSGIPCELGELTFLFLAMNFSYNMLEGPIPGTQIQSONSSFAENPG
LCGAPLAKCGGEEEDKEEKDKGLSWAAAIQYVFLGFLAIGHILTSYKRDWF
MRIFSCFSSPLK"
1753..3194
/gene="dl3000w"
/number=1
3195..3260
/gene="dl3000w"
/number=1
3261..4428
/gene="dl3000w"
/number=2
complement(5905..6149)
/number=1
5905..7913
/gene="dl3000w"
complement(join(5905..6149,6243..6344,6436..6826,
7236..7913))
/note="strong similarity to glycine
hydroxymethyltransferase (EC 2.1.2.1); contains EST
qb:1565775"
/codon_start=1
/product="hydroxymethyltransferase"
/protein_id="CAB10172.1"
/db_xref="GI:2244749"
/db_xref="SPTREMBL:O23254"
/translators="MEPVSSMGNTSLVSDPEIHDLEKRRKRCQRIELIASENFTS
FAVTEALGSALTNRKYSIGIPGNRYGNEFIDEIENLCRSRALEAFHCDPAAGVNVQ
PYSGPANFAAYTALQPHDRIMGLDLPSSGHLTHGYTSGGKTSATSIYFESLPYK
VNFTGVIDYDKLEKALDRPKLLICGGSAYPRMDYARCPAIAKVKALLCDMAH
LSGLVAAQEARANPEECVDVYTTTHKSLRGPAGMLFYKKGPKPKKQPGGAYDDE
DKINAFVPAQQGPHHQIGALAVALKOANTPGFYTKAQYKANAVAGNLTMSKGT
QIVTINGHLVLDWRDLPUGLSTGNRKVKELCDLCSITLNKNVAVGDSALAPGGVRIGA
PAMTSRGLVKEDEQIGFELSRATVLTLDIQTYTKLLKDFNKGVLNNKDLQDKADV
EKFSASYEMFCFLMSMKYKD"
complement(6150..6242)
/number=1
complement(6243..6344)
/number=2
complement(6345..6435)
/number=2
complement(6436..6826)
/number=3
complement(6827..7235)
/number=3
complement(7236..7913)
/number=4
12823..14568
/gene="dl3010w"
join(12823..13533,13822..14568)
/note="strong similarity to adenosylhomocysteinase (EC
3.3.1.1)"
/codon_start=1

```

```

/product="adenosylhomocysteinase"
/protein_id="CAB10173.1"
/db_xref="GI:2244750"
/db_xref="SWISS-PROT:O23255"
/translators="MALLVEKTSGGREYKVKQMSQADFGRLLELEAEVEMPGLMACRT
BFGSPQPKGARITGSILHMTIQTAVLETLTALGAEVRCNSNIFSTODHAAALARD
SAAPVAMKGETLOEYVWCTERALDNGGPGGPDLLVDGGDALLIHGKVAFEIPEKT
QGVDPDSTNDPERQIVLSIIEKGLQVDPKKYHKHMERLVGYSEETTVGYKRLQMQQ
NCTLLFAINVDSVITKSFEDNLYGCRHSLDPLDMRATDVMIAGVAVICGCGVDGKG
CAAMKTAGARVIVTEIDIPICALQALMEGLVLTLEDVVSEADIVFTTGNKDIIMVD
HMRKNNAICATGPHSVFMSCTFNQVLAQELMNEKASRYEKYKVVLPKRLDEKVAL
KRLNMLCATGPHSVFMSCTFNQVLAQELMNEKASRYEKYKVVLPKRLDEKVAL
HLGLGARLTKLSKDSQSDYSVIDIEGYPKPHYRY"
12823..13533
/gene="dl3010w"
/number=1
13534..13821
/gene="dl3010w"
/number=1
13822..14568
/gene="dl3010w"
/number=2
complement(16093..16270)
/number=1
16093..16501
/gene="dl3015c"
complement(join(16093..16270,16317..16501))
/note="weak similarity to Walleye dermal sarcoma virus gag
polyprotein"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10174.1"
/db_xref="GI:2244751"
/db_xref="SPTREMBL:O23256"
/translators="MFNSTALVIFAILFLISADAFPIPSNGEIDAMLIRNSIIGED
EDLPATISRRVLAQRNOELLTIVALKLITAEAVTPLPDALETQTSTHBEKI
QRIAKIGSFIAEASKLYD"
complement(16271..16316)
/number=1
complement(16317..16501)
/number=2
complement(join(18573..18914,18997..19137,19212..19554,
19603..19719,19762..20123))
/note="weak similarity to fusion protein 1a/1b - sugar
beet"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10175.1"
/db_xref="GI:2244752"
/db_xref="SPTREMBL:O23257"
/translators="MDHVSSLPEVLYHILSFTTKEAALTSLSKRWRNLFTFVPNL
DIDDSVFLPHQEGEDREYETOKFMKFDVRVALQNSPIKLSLKLKRGFDSHRVDG
WISNALRGVTELDLLILMLVTLKNSVRVDWLAAGDIFLPMKLTVLHSLVRLVDN
VNWGKQVNVSNASIKITVNYNICLTGVSVDTPSLVFCFTEYVAIDYPPVIMQNLF
EARSLSVTQDIERARAPNIDWEDVLRVLNMEKLIKIRNVQYMYLSLTVLEVL
SLCCSPMPVNNLKSITLKSRSRQWQMPVLLRNCPLHPTLVLRGILHHVTDKGYA
CDVSRREKGGSLTSCPVKLEIKGPGCTIKEMHMKHFVNYPCLMETNIIIESGP
PKLRVPQVCEFTAEEMHEHYKSSSFNVQLLVGGYLYKKWTAQ"
complement(18573..18914)
/number=1
18573..20123
/gene="dl3020c"
complement(18915..18996)
/number=1
complement(18997..19137)
/number=2
complement(19138..19211)
/number=2
complement(19212..19554)
/number=3
complement(19555..19602)
/number=3
complement(19603..19719)

```

```
/number=4
complement(19720..19761)
/number=4
complement(19762..20123)
/number=5
20514..20606
/note="EM_ORG:CHSOPSBO Spinach plastid psbB operon with
Query Match 19.9%; Score 300; DB 7; Length 200576;
Best Local Similarity 55.7%; Pred. No. 2.5e-55;
Matches 768; Conservative 0; Mismatches 530; Indels 80; Gaps 7;
QY 36 tgcagcaagcttctcgcgaagccacataatcccgccctcccaattccccaagagac 95
Db 81653 TGCCTTGAACATCCCAAGCAAGGTACATCAACCGCGCTTCAACTAGCAACCGCC 81594
QY 96 tcctaaaaacggcagctgaagtcacatttttcacgagaggtttatgcattgagcgcatgg 155
Db 81593 TCATCCACCGGCAACCGTCCATCTATCCACCGCAGCTCTCTCACCAGCATGG 81534
QY 156 ccaacacagctccgcgcgtccggaaacccaccggccgcgcgactctgtgctgtctccg 215
Db 81533 GCGAGCCACCTTCCA-----CAAAAGGTCTATCTCTTCTGCTTGTTCACCG 81489
QY 216 acgctacgagcagggctgaagccctggcgagcgagggagcgcgtacatgtcccgatga 275
Db 81488 ATGGATTGCGACGCGGTCTCACTCATCTCGAAGCCAGCAAAATCTACATGTCCGAATCA 81429
QY 276 aagcccgcggtccgagcgcttaagaaacctctctctcaacaaccacgacg----- 326
Db 81428 AACGATGTGGTTCAAACGCGCTGAGAGACATCATAAAGCCCAATCTTGACGCCACCACCG 81369
QY 327 -----tcagttcgtctcgtctcgtccacctcttctgcatggcgcgagatgg 374
Db 81368 AACAGAGCCCTATCACC GGGTAACTACTCTGTCTGCTCCGTGGGTCTTACGGTAG 81309
QY 375 cgcgtgagtcacaggtcccgagcgccctctctgtggtgagcgcccgccacctgtgtgca 434
Db 81308 CGCGTGAGTTTCACTCCCACTACACTCTCTCTGGATTGAACCACTACTTACTAGACA 81249
QY 435 tatattcttactcaacgcgtacgcagagagatgacgcgggttccgacgaattc 494
Db 81248 TCTACTACTACTACTTCAACACCTCTTCAACACATCTTTCACGTTGAACCGATTAAAT 81189
QY 495 agctccctcgggttccacctggagcagcgagcttctccgaccttctgtgtcgcgaga 554
Db 81188 TACCGAACTGCCACTGATCACCACCGCGTGACCTCCCTCGTGTCTTCAACCTTCGAAG 81129
QY 555 caccggagagattccggttgatgataagagagagctgaaactttagcggtagaaga 614
Db 81128 CATTAACGCTACGCTTTGTGACTCTAAGAGACATATCGAAGCTCTCAACGGAATCAA 81069
QY 615 aggcgaagattgttgtaacacgtttgatcggttgagcccgatgacactcagcgtattg 674
Db 81068 ACCCTAAGATTCTTTGAACACATCTCTGCTTTGGAACACGATGCTTTAACTCTGTTG 81009
QY 675 ataggtatgattgacggatcgcccttgaattccctccgaccttcttgacgcgcgag 734
Db 81008 AGAAACTCAAGATGATCCCAATCGGACCGTGTGTTCTTCTCCGAGGGTAA----- 80955
QY 735 atccctccgaaaagtcttacggcgccgatcttttcgaaaaatcggaggaggaataactgcg 794
Db 80956 -----AACCGATCTTTTCAAAATCTTCCGACGAGGAT---TACA 80922
QY 795 tggagttgttgacacgaacgcaaatcttcggtggtgatgtcgtttggga---gcg 851
Db 80921 CGAATGGTTAGACTCGAAGCTCGAGAGATCAGTGAATTTACATTTCTTAGGCACACAG 80862
QY 852 ttttgaggtttccaaagccaaatgaagagattggaaaggggctattagcctgcgaa 911
Db 80861 CCGATGATTATCCAGAGAAACACATGGAACGCGTTACTCACGCGCTGTAGTCAAAACA 80802
QY 912 ggcggttttatgtagatcagagacagagaaatgacgcgcgcgagcgcgcgaagaagaag 971
```

```
Db 80801 CACCGTTTTTATGATCGTCTGAGGAGCAAAAATCCAGAAG-----AGAAAGAAAGAT 80749
QY 972 agttgagttgcattggggaattgaaaaaaatgggga-aaatagttcttcgtctgcag 1030
Db 80748 CGGTTTCTGAATTTATGATCAGAGGAAGTATCGAGGATTTGTTGGTGGATGCTTCTCAG 80689
QY 1031 ttgagatttcggcaccctcgcttgggatgttttcgtagcagcatttggtggaactcg 1090
Db 80688 ACAGCTGTTTTGGCGCATTTGCTGTGGGATGTTTTGTGACTCATTTGTTGGTGAATCG 80629
QY 1091 gctgtgagagcttgagttgcggggttcggggtggtggcggtgcccgcagtggtgatcag 1150
Db 80628 ACGTTGGAGAGTTTAGAGAGTGGTGTCCGGTGGTTCGCTTTCCGACGTTGCTGATCAG 80569
QY 1151 acgacgaatcgcaagctgattgagatcgctgggggacagcgggtgagagtgaatgaat 1210
Db 80568 TGTACAACGGCAAGCTTTGTGGAGGATACGTGAGGATTGGAGTGAAGTGAAGTTGGG 80509
QY 1211 gaaagggtgggttgatgactgagatagagaggtgctgagagatggtgatgagtggtg 1270
Db 80508 GAGGAAGCAGATGTGGATGGGAGGATTAGAAGGTGTTTGGAGAAGTGTATGAGTGGT 80449
QY 1271 ggtgagaagacaactagtagagaagaatgccataaaatgaaactttggccagagaa 1330
Db 80448 GGAAAGAGCGGAGGAGATGAGAGAAATGAGAGAGTGAAGAGCGGATGCTGTGAT 80389
QY 1331 gccatggagaggtggtcttccactcaagaatctcaacgccttcttcacaaagttg 1388
Db 80388 CGCGCAGCGGAAGTGGACGCTCGGATTTGAATCTTTAAAGTTTTGTGGACGAGGATG 80331

RESULT 7
AB000623 LOCUS AB000623 1594 bp mRNA PLN 05-FEB-1999
DEFINITION Nicotiana tabacum mRNA for glucosyl transferase, complete cds.
ACCESSION AB000623
VERSION AB000623.1 GI:1805358
KEYWORDS JIGT; glucosyl transferase.
SOURCE Nicotiana tabacum (strain:BY-2) suspension culture CDNA to mRNA.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1594)
AUTHORS Kojima,H.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1997) to the DDBJ/EMBL/GenBank databases, Hlsae
Kojima, school of agriculture,nagoya university, laboratory of
biochemistry; nagoya, furo-cho,chikusa 464-01,japan
(P-mail:hlsae@nuegri.agr.nagoya-u.ac.jp, Tel:052-789-4098,
Fax:052-789-4094)
REFERENCE 2 (bases 1 to 1594)
AUTHORS Kojima,H., Hashizume,K., Imanishi,S. and Nakamura,K.
TITLE Jasmonate-induced potential glucosyltransferase from tobacco
suspension cell
JOURNAL Unpublished (1997)
FEATURES
source
1..1594
/organism="Nicotiana tabacum"
/strain="BY-2"
/db_xref="taxon:4097"
/tissue_type="suspension culture"
52..1455
/gene="JIGT"
52..1455
/gene="JIGT"
/standard_name="jasmonate-induced potential
glucosyltransferase"
/codon_start=1
/product="glucosyl transferase"
/protein_id="BAA19155.1"
/db_xref="GI:1805359"
```

```

/translation="MENLKNECHVLIALPQGHINPSLOFSKKLLINLCVKVTLSSSL
SAFNRINLKPKEGLTFAPESDGYDGNFKGSDDDHLFNSAIFKSHGSEFIANLRSKA
KNGPFTTIVITILMDWAGSVAKKLHPISTLFWIQDPAIVFDIYYRFTFANYFNKND
SODQIIEPLGPSLSSDFSFDDVKNDMAVESIKROIELSENSEYNNPWLNTFED
ALELNAVLKNTVMGIGLIPSLFDEKDRKDNFFRADMIENSENMYEWLDAANK
VITVAGSAVETASSOMERISOLLLKCGRPFLWIRETLNGKEPEKLTCKDELEKI
GRVWKSOMEVLKHSVGCFLTHCGWNSTLFSLASGVPVACPIWQDQICNAKLQD
VVKIGVRVNAKEGLIKRDEFQKCIIEIVGDAEECEELKRNQAKKKDLAKESTKENS
SVNKLKAYVNECLLGHIC"
polya_site
BASE COUNT 514 a 238 c 325 g 517 t
ORIGIN
Query Match 18.8%; Score 283; DB 7; Length 1594;
Best Local Similarity 60.9%; Pred. No. 1.1e-51;
Matches 501; Conservative 0; Mismatches 310; Indels 12; Gaps 2:
QY 578 atgaagagaagctggaactttagacggtgaagagaagcgaaagtgttgggtgaacacg 637
D 540 ATTAAGGCAAAATGAGATATTGATAGTGAAGAAAATCCAGAAATCTTGTGAATACT
QY 638 ttgatgccttagaccgagcactcagcgtattataggtatgagttgacggatc 697
D 700 TTTGATGCATTGACTTAATGCTTTGAGGTTTGTGAAAATGTAATACTATGGTGGAAAT 759
QY 698 gggccgttgatccctccgcctctttggacgcgagagatccctccgaaacgtcttcacgc 757
D 760 GGACCTTAATTCCTTCATCTTTCTTCTGATGAGAAAGATCGAAAGATAATATTTTCGCG 819
QY 758 ggcgatctttcgaaaaatcggagagagataactgcgtgagtggttgacacgaagcg 817
D 820 CTTGATATGATGAAA-----GTGAAATAATTCATGAGAGTGTGGATGCAAGGCG 873
QY 818 aaatcttcggtggtatgtcgttctgtggcgctttttgaggtttcccaaggcacaatg 877
D 874 AATAAATACGATTATTACATACATATTTGCTAGCTATCTGAGATATCAAGTCAATGGATG 933
QY 878 gaagagatgggaagggcctattagcctgcgaagggcgtttttatggtgacacagaa 937
D 934 GAAGAGATTAGTAGGATTGTTGAAATGTGGAAGACCAATTTTGTGGGTGATTAGAGAA 993
QY 938 cagaagaatgacgacgagcgaagaagaagaagatgagtgatcgatgggggaatgaaa 997
D 994 ACCCTAAAT-----GGGGAATAACCCGAGGAAAAGTTGACTTTGTAAGAGATGAATTGGAA 1047
QY 998 aaaaatgggaataatagtttcgtgctgcgagttgagaggtcttcgcccacccctgcgtg 1057
D 1048 AACATAGGAGAAATAGTCGCTGTGCTCGCAATGGAATGGAATGCTTAAACACTCGTCCGTA 1107
QY 1058 ggaatgttcgacgacattggttggaactcggctgctgagagcgttaggttcgggggtt 1117
D 1108 GGCTGTTTCTTACTCATTTGGATGGAATTCGACTCTGGAGAGCTTAGTCTTGCTGTG 1167
QY 1118 ccggtggtgagcgttcacagtggtttgatcagacgacgaatcgagagcgttagtgagat 1177
D 1168 CCATCGTGGCATGCCGATTTGGAATGATCAATTTTCAACGGAAGCTCAITCAAGAT 1227
QY 1178 gctgggggacagggggtgagagtgagaatgaagggggtgggttgatgagatcgag 1237
D 1228 GTTTGGAAGATTGGTGTGAGAGTGAACGCTAATAAAGAGATATCAATTAAGAGACGAG 1287
QY 1238 atagagagtgatgagatgagtgatgagtgagtgagtgagtgagtgagtgagtgag 1297
D 1288 TTCCAGAGTGTATGAGATCGTAATGGAGATGCTGAGGAGGAGGAGGAGGAGGAGGAG 1347
QY 1298 aatgcataaaatggaagactttggccgagagaagccatgggagagagatgagatctcact 1357
D 1348 AATGCCCAAAATGGAAGATTGGCTAAGGAATCTACGAAGGAAAATAGTTCATCTAAT 1407
QY 1358 aagaatctcaagcctcttctcaatcaagtgcagtgcttaat 1400

```

```

Db 1408 GTAAATCTTAAGCCTTATGCTTAATGAATGTTTACTTGCTCAT 1450

RESULT 8
AC005106/c DNA PLN 16-DEC-1999
LOCUS Genomic sequence for Arabidopsis thaliana BAC T25N20 from
DEFINITION Chromosome 1, complete sequence.
ACCESSION AC005106
VERSION AC005106.2 GI:6587719
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 84203)
AUTHORS Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H.,
Conway, A., Conway, A., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M.,
Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and
Ecker, J. R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T25N20 from
Chromosome 1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84203)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 84203)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 84203)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 5 (bases 1 to 84203)
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C.,
Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chio, J., Choi, E.,
Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T.,
Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S.,
Mukharshy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J.,
Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R.,
Federspiel, N., Theologis, A. and Ecker, J.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT On Dec 16, 1999 this sequence version replaced gi:3366536.
FEATURES
source
Location/Qualifiers
1..84203
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(join(1491..2159,2389..2452,2568..2703,
3051..3149,3231..3277,3369..3457,4022..4139,4246..4353))
/note="5' end of hypothetical protein YP8H12.1; overlaps
37 end of YP8H12"
/codon_start=1
/evidence=not_experimental
/product="T25N20.1"
/protein_id="AAC80580.1"
/db_xref="GI:3935137"
/translation="MGKKEQDHSVVEVDKVEAVLHLLRKHSLTLTKQKFCNRC
VGRFLRIKGDVKKAAQLRSLRSLSLIEADTEFLAEGLAYVAGLDECR
PVLVFRIKQDYQKLTKQLTRLVVFLVAISTMSRNVEQFVFLDAAAAVLLVTL

```

CDS
IVLFESLNVPLLLILCNSTAHLPPFFFFFIYKTNPKLLQISIFHEHTCDYTONSRRG
IRFTVDSLTAMIVSQDDQDFDYDDSSSPRSVSLRFDTSLSLKSTDKIGSCASS
RFAFTVSDGLDTVPCWLTLDYDTSSTKLGHNTGAYISPLNARSFSPASPAARSEFG
GPRSFASFPARTDRHSIGTLDRDIPRPSPFQSPAFIFRRSHVSKSEKPRDS
VQFLAFYRPDYEMTYRKSMPPLGGVLVSIVSQIRRHVSLSQRP"
Join(9304..11268,11387..11513,11607..11692,12103..12341,
12698..12784,13257..13356,13451..14263)
/note="unknown; similar to ESTs dbj|C26511 and
gb|A105450"
/codon_start=1
/evidence=not_experimental
/product="T25N20.2"
/protein_id="AAC80581.1"
/db_xref="GI:3935138"
/translation="MEGGGSGERSIRIASNPSRLSKKSDQGVRTVSSTETRONYOKR
SRMVSDSSDEPNKPPRRSGVDRKTLGAKEKPVKRORVREDNRGVRRNEASG
SEMKMKLIDIFEDVDCPDSANLMEKPFDSVGRSFSFASRRVDSYVSGSGSR
EGLFDRRTFVNGCSASSQDSDSDPEMRVQGVINGVLYKYNNKNTLAAISI
NPDAEIIYRPSSQAQRRENVVYKPSNNVNNSEESDMSRKKRKKSEYS
KPKAEYTKSTFFPELVNPDVREERRGGTDQRLREIRIKGMLTDAGWTIDYKP
RRQSTLDVYVNPSTAYWSIIKAYDALLKQDKDEYDARPKRDAAVASVEEIVN
KLAKAKTRSEMTRKWNQSGSDSENKSEGAYTDTSEIRISRTKLGGCKTKKGR
NGADWELHKKSRLSYNNARPSCGSDSHYLHGKTKKIGRCTLLVRSSCKKNPAI
NGPNPYSKARTLLSLIESGVVQLRQVYMRRGKVMLEGWTTREGIHCDCCKIL
TVSRFTHAGSKSCOPFONILYESGASLLQCVRAWNNKDATNLAHOVDTGDGPN
DDAGCICGGGDLICDCGPTSYHONCLGMQVLPDGMWCHPCNCKCDAVASGKD
GNTISLSGMRERYHQCLNDEAHKYVSGSASSCCPKLEFLFKLQKYLGVKTE
IEGYSWSLHVDDTDSNMQSAQRIENSKLAVLAIMDECFLPIVDRRSVDLI
RNYLYNCGSNFRINRTGYTALLERGEDEIIISAASLRFHGHQLAEMPFICTHRIYRQ
GMRRLDIAESAMRSKVELVIPAIPDFLHWTGNFGFTPLDSDVRKEMRSLNIV
FGIDMLQPLHEENIAPAAGAMASIVETETKSEFTSSVEIGPYAVEGDEFVAD
AANCYDILASDEDNLIVAVETAMGTICKPKDELSRHFRGEESGISSPCQI1LKSGT
KVLVGHCDTGGSCDGLTDVNVPADASLLSQEQOASAFKVENNLSLSISGRSS
DLSISQEVKRSQTSNLDGVPCKDYNLLVPGAKLKDSDADAPGFLI".15721))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T25N20.3"
/protein_id="AAC80582.1"
/db_xref="GI:3935139"
/translation="MGFLVAVNFSPTLVHHMKSQPCQNEKLRQOQTSSLDRRGF
LKCVGSGFMATIEFSGAQASSEKLODEGVVYGAFTLFDPNETKSGKEPLRAY
LKSAREVMTRESUKENPKNAKERRSADAKEISRDYLSNWRGQTVAGESEYVEL
ENVIRALAFYSKAGSAPLPDVEKTEILDNLKAEFL"
Join(16409..16661,16979..17133,17234..17258,17346..17362)
/note="unknown; similar to EST emb|226799"
/codon_start=1
/evidence=not_experimental
/product="T25N20.4"
/protein_id="AAC80583.1"
/db_xref="GI:3935140"
/translation="MMMDPPKFEALAPVYVYVDMKRCVPDDYDARRVGPCIKRIL
KISGVNPTITAVGSLKVPDRDILEVYSSGTISLYHEVARPLEEEDSCSETPGPAS
WGSVCRRICGFIAGQGVNDPITHLSTRHEHLKRQFTNPKGHLAN"
complement(Join(17577..18026,18231..18796,18981..19410))
/note="similar to RNA polymerase II elongation factor
homolog (S59773)"
/codon_start=1
/evidence=not_experimental
/product="T25N20.5"
/protein_id="AAC80584.1"
/db_xref="GI:3935141"
/translation="MFLHVLVLEAASIGFVIECRSYNNDSMDIDQSNIGESPHLLLRP
VSPLESGEPLTAPENPNPGDTMHWKPRISGKGVFVDRYLPYKQVLPGLDTEILR
KNVFRSLSLORYIRVHPEADVOKFASFSWSIPCRDGGVLPKQVLPVYSDE
DPMRDGDTAPVCKACNEKCRSLMPQCEATLPAWPCDICGGEKFCVDCCLLCKL
ISLEHGGYSYIKCAEVYSBGHGVHNMALRAYLAGTIGGSKGLDYLRGGRDA
KDLFPHVNFLEICQVYEQGVDEKILNLTGICILRGAGRDNAKELNCTEYVIRII
FILVYDSGSGARENDTLQSLQDTPIGPIPFHNEAHMKLEEEIEGVLARLKAQEV
FETQIAEGKLHQAQKLSIQLEKRSLSRVSSTGANSLSMTNLYLKLRLDQIRKEV
TKLKEEYVAKGFRTPRGVLEYPHLNIED"
22098..22826

CDS
CDS
CDS
CDS

/note="unknown; similar to ESTs emb|233701 and
emb|234709"
/codon_start=1
/evidence=not_experimental
/product="T25N20.6"
/protein_id="AAC80585.1"
/db_xref="GI:3935142"
/translation="MUNTSLSLFLTKTKTPRYMKNFHLCPFSNLTUKSSSPSGAT
SDPNRPSILLNNFLLYDDSSAAHRLSKPLIHDVPSSTFTASTSTAAANSSSA
SYDDSDNFGAPDDSDPDLTAVLASRRFPSPGCSNSITSPDRLCRDNDYATR
LLTGCTTAVKHVQSPDPYNDFRRSRMOEMIDAVTNAGDLRRYEFLEHLLSYLSLNAAD
THKFIITRAFADILVLSLSDGHRIS"
complement(Join(26300..26503,26613..26880,27004..27259,
27394..27499,27748..27804))
/note="unknown; similar to ESTs gb|T76155 and gb|N96012"
/codon_start=1
/evidence=not_experimental
/product="T25N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAMAALNQLFVVLASKPQEKITPEESRAIVSFHFKALMTAGFA
SGVGGGLTMQVTKKLLPKGLERVALAGNYLHIFVYDHSFLRSEKCVFFSEFTSQ
ELLRLILLGIGQVPMQFPHLIITSLVMLHECRKSLVRSNRGEAWRQMLSKHPY
SESVYGDGDKPQMRWRRTTETELASSYDDVNATKSORNPGLPNPSHRRISSGSDA
SKYKOTLONSGNSGEMAREVDLIVPGCSATESIPAPVLSKLASKTQTRQKRAQ
RRGLANREASTNTPQYELA"
complement(Join(28447..28740,28976..29125,29206..29943))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T25N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translation="MKSRAENQMINRETMGKILVSPSEKSWTSSTTTTTTTTTVS
TEWDRSDNCFYPCDKRANCSCIEDLSNATLDLMLPSVQKSSUTKLSSASNFKS
TVESPTSFDPYVTPPASVSRPILKMLSESPKLLTKKSKVFENEQRTTKERRL
LIVYFLKILVIGLVGLBELGFSWVLEEVKPEFTTEIVRNAYEQATDLGAVKRL
LEDEFKGFVTSRKSCTGSDSKINODGPMKSKVLKYSATEEYSICWNPLOTQY
LFHTGSSSITVTLSCGRVTEGKFGYSIHETNTSWKTKWTSVLQDPNTWVLEY
SQSSVMDSSLSLLSDIMEHLVFRVAKNNRERYWFFSSSRRLYREAESKASATP
T"
Join(32077..32407,32715..32956)
/note="similar to lipid transfer protein (AB007843)"
/codon_start=1
/evidence=not_experimental
/product="T25N20.9"
/protein_id="AAC80588.1"
/db_xref="GI:3935145"
/translation="MNSNLSFLSALIFSLSSNSPTSLAQINTPCSPSMLSSVTGC
TSLTGGSPSTSDCCGALKSLTGTGMBCLCLIVTAGVPIISIPINRPLAISLPACGI
PGVPVQCKGPASGPTSPDSQSDPGETRELFTATSSNSKTCFSEFVSKRFSGLL
TSNHSGLLUSVRLPLQRKLLMVNHNIS"
complement(33000..33155)
/note="similar to EST emb|226808"
/codon_start=1
/evidence=not_experimental
/product="T25N20.10"
/protein_id="AAC80589.1"
/db_xref="GI:3935146"
Query Match 14.8%; Score 222.8; DB 49; Length 84203;
Best Local Similarity 53.4%; Pred. No. 1.7e-38;
Matches 750; Conservative 0; Mismatches 567; Indels 87; Gaps 10;
QY 13 aaaaatggctccgcgcgcgctgctgctagcaagcttccctgcgcgaagcccaataatcc 72
|||||
Db 69940 AANAATGGCGCAACCGCATTTCTACTGTAGCTTCCGGCGCAAGGTACGTGNACCC 69881
|||||
QY 73 cgccctccaatcgccaagagactccctaaagcc---ggcactgacgtcacattttcac 129
|||||
Db 69880 ATCTCTCGTTTTCGTCGTCGGCTCATCAAAACAACGGCGCACGTGTAACTTTCCCCAC 69821
|||||
QY 130 gagcgtttatgagcgccgcacagcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 189
|||||

FEATURES	source	'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://compio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCAN.html), Fexa (V. Solovyev & A. Salamon, Sanger Centre, http://genomic.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).	
		Location/Qualifiers	
misc_feature	1..33654	/organism="Arabidopsis thaliana"	
		/cultivar="Columbia"	
gene	CDS	/db_xref="taxon:3702"	
		/chromosome="1"	
misc_feature	1..33654	/note="Overlap with bases 50550-84203 of 'RAMU' clone T25N20, gb AC005106. Features from bases 1 to 29,000 are annotated on overlapping clone T25N20, gb AC005106."	
		complement(29404..40201)	
gene	CDS	/gene="F3P20.1"	
		complement(join(29404..29652,29751..29936,30023..30073,30150..30950,31055..31284,31428..31486,31568..31651,31740..31858,31981..32100,32252..32411,32506..32695,32780..32848,32961..33086,33172..33281,33357..33508,33634..33700,33851..33948,34038..34161,34243..34403,34491..34575,35347..35441,35543..35680,35806..35892,35975..36034,36112..36220,36644..36981,37155..37272,37354..37473,37725..37818,37915..37991,38079..38204,38319..38445,38591..38682,38787..38879,38964..39080,39165..39245,39382..39505,39608..39729,39823..39923,40015..40201).)	
misc_feature	/gene="F3P20.1"	/note="Highly similar to putative callose synthase catalytic subunit; Highly similar to cotton putative callose synthase catalytic subunit, gb AAD25952."	
		/codon_start=1	
gene	CDS	/protein_id="AAD30609.1"	
		/db_xref="GI:4836907"	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	AGKQSDAREMSQFYQHYKKYIQALLNADKADRAQLTKAYQAAVFEVLKAYNOT	
		EDVEADENLVDDILDIITQIYVFNILPDPSQQAIMRLPEIQAAVAALRNQPL	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	LPMTAGHKKLDDILDLQSMFGKONVNLNREHLNLLANVHIROFPKPOQKPL	
		DRALTIYMKFLRNFKYKWKYLGKSLWLTQIQEQQRKLYMGLYLLIINGERAN	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	LRFMPCLCYIYHLYGMLAGSVSPMTEGKVPKPYGDEDAFLQKVTPYIQTISKEA	
		KRSRGKSKSHVNRNDDNENYFWSIRCFRLGMPMRADDFCOTAEELRSEITKS	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	NSGDRMGKVNVEIRSPWHIFRSPDLWSFVILCLOAMIVIANWGSGLSAIFQGDV	
		FLKLVSLTAALIKAOALDIALSWKARHSMYLYVKLVYMKVGAAYVVMVAVT	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	YASWKNASGFSOTIKNFGSHNSPLFVAILIILSPNLSALLELPFFIRVYLE	
		RSDIKMLMWSQSYFSPSQNLPLVQPKDQIMRTHISVISWHEFFPHANKNGV	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	DLVSPNVLVPMDFQIWAIVSLVGLNGAFRRLEIRITGLMRSRQSIPIGAFN	
		DIALSPNDSOTKKRRFATFRKFDLPSSKDEAARFAQMNKIISSREEDLSD	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	REMLLYLVSPDLDLRWPPFLASKIPTALDMKSDKGRLEKRLKRLAVDSYNT	
		CAVREYAGKPNLYLVVGERGVINDIFSKIDEKETLITELNLSALPDLYGQ	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	FKVLEYLLENREEDKQIVLVLLANLELVRDIMEEVPASNTSVNFDSQFLIKRLK	
		GKQKQIKHLHLITVESAMDPVNSLEARRLITFESLENFMPAPKIKNMLSPSVL	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	TPFSEDVLFISGLCEQNEQDGSILFYQKIFPDENFTNFMVCKGNEELRAREDL	
		EELRLWASTRGOTLTKVGMAYYRKALEQAFLOAKDRAKDLRLMTTPYIRVAYIDE	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	KSGSLWAQALADMKFTFVSCQYSIHKRSGDRAKDLRLMTTPYIRVAYIDE	
		VEQTHREKYTEKEIYYIYALYKAAQPKMDSSESQVTLQIYIKRLPGPALIEG	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	KPNQNHAIITFREGLOITDMNQNMEEAFKRNLIQEFLEKGGVRCPTILGRE	
		HIFTGYSVSSLANQNSQNSFVITIGORVLAAPLVKRVHYHPDIFDLFLHTRGFNST	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	LRGENTHREYIQVGRDGLNQISMEPAKITANGRDOTLSRDLYRLGHTRGFNML	
		SCYVITIGVFSMLPLVYVYLYGLYLVLSGLEGLSSORAFRNKPLEALASQ	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	SFVQIGELMALPMMEIGLERFHNALEFVLMLQLOLASVFTFQLGKTHYQRTILF	
		HGAQYRGTGRGVVPHAKFAENRFRYSRHPVKGIELMLLLVYIQFGSGVGVTY	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	ILITVITWFMVWMLFAPLPNPSFGWQKIVDDMTDNKWIYNRGGVYPPGKSWFI	
		WKEKLEHRSGVGRITLIEFLALRFFIYQYGLVHLSTFKGNQSFVYAGSWFVI	
misc_feature	/translation="MAOREPDPPPORRILRTOTVGSIGEALDSEVPVSLVETAPILRAVEASNPRAVAYLCRFVAFKAHRLDPTSGRGVROFTALLQRLERSEITL	LFILLVIGLGVGRFRFNLQIFRIKGLVFLFVAILITFLALPLITIKDITCM	
		LAFMPTGWGMLLIAQACKPLIQOLGIWSSVRLTARGYIVMGLLFTPVAFWAFVY	

Query Match

14.8%; Score 222.8; DB 8; Length 103223;

Best Local Similarity 53.4%, Pred. No. 1.7e-38;
Matches 750; Conservative 0; Mismatches 567; Indels 87; Gaps 10;

Qy	13	aaaaatggtccgcgcgcgtgctgctagcaacgttttcctgcgcaaggccacataaatcc	72
Db	19391	AAAAATGGCGCAACCGCATTTTCTACTGGTAACGTTTCCGGCGCAAGGTCAGTGAACC	19332
Qy	73	cgccctccaaatctgcgcaagagacactctaaagcc--ggcactgaagctcacatctttcac	129
Db	19331	ATCTCTCCGTTTGTCTGCTGCGCTCATCAAAACAACTGGCGCACGTTGTAACTTTGGCCAC	19272
Qy	130	gagcgtttatgatggcgcgcgatggcccaacacagcctccgcgctgcgcgaaacccacc	189
Db	19271	GTGTCTCTGTCTATTCACCGCTCTATGATCCCAAAACACAAACGTCGMAAAT----	19215
Qy	190	gggcctcgactcgtggcggtctccgacgctacgacgagcggtctgaagccctggcgga	249
Db	19216	----CTCTCTTTCTTACTTCTCGCGAGATTCGAGCGGAGTCACTCTCAACACCGCA	19161
Qy	250	cggaagcgtcacatgtccg---agatgaagccgcgcgtccgagcgtcttaagaacct	306
Db	19160	CGACGTCCTCAAAACCGGTTGGTACACTTCCGAACGTAATGGCGATAAAAGCTCTATCGGATTT	19101
Qy	307	ctctctcaaacaccgca-----cgtaacgttgcgtctactcccaactctt	354
Db	19100	CATCGAAGCTATACGATGGTGACTCTCCGCTAAGTTGTTCTGATCTACAGGATCTTCC	19041
Qy	355	tgcattgggcgcgcgcgtgagtcacaggtcccgagcgcctctctctggtcga	414
Db	19040	CAACTGGGTTCCAAAGTGGCGCTAGATTTTCACTTCCCTCTGTTCACTCTGGATCCA	18981
Qy	415	gcgcgcacgcgtgctgctgcataattactctactcaacgcgtacgacagagatcga	474
Db	18980	ACCAGCCTTCGCTTTCGACATTTATTACAATTTACTCTACAGG-----AAACAA	18933
Qy	475	cgcggttccgagaaattcagctccctcggcttccaccctggagcagcagctctcc	534
Db	18932	CTCCGTTTTCGAGTTCCCGAATCTACCTCTCTCGAAATCCGCGATCTCG-----CTTCT	18878
Qy	535	gaccttctgctgcgcgagacacgcgagagattccggtgatgatgaagagagactgga	594
Db	18877	TTCTCTCACCTTCCAAACAGCAACCAAGCCACAAACAGTATATATCAA--GAACGTATGCA	18819
Qy	595	aactttagacgctgaagagaagcgcaaaagtcttgtaaacacgtttgatcgcttgagcc	654
Db	18818	TTTCTCAAGAAGAATCTTAACCCGAAATTTCTCGTCAACACATTCGATTCGCTGGAGCC	18759
Qy	655	cgatgcactcagcgtattgatgataagttgatcggatcgggcgcgttgattccctc	714
Db	18758	AGAGTTCTTAACAGCTATTTCGGAATATAGAAATGTGCGAGTTGGTCTTTACTTCTCTGC	18699
Qy	715	cgctctctggacgcgagatccctccgaacgtcttaacggcgagatcttttcgaaaa	774
Db	18698	GGAGATTTTCACTGGGAAGCAATCAGGTAAGATTTATCAAGAGATCATC-----	18647
Qy	775	atcgaggagaaataactcgcgtgagttggtgacacgaagccgaaatcttcggtggtga	834
Db	18648	----AAAGTAGTAGTTATACACTTTGTTTAGACTCCGAAACAGAGTCTCTGTTATTTA	18594
Qy	835	tgtgcgtttggagcgcgttttgagtttccaaagacacaaatggagagatgaggaaagg	894
Db	18593	TGTTCTTTTGGAACCAATGGTTGAGTTCTCGAAGAAACAGATAGAGAACTACGAGAGC	18534
Qy	895	gctattagcctcgcgaagccgctttttatgatatcacgaaacagaagaatgacgacgg	954
Db	18533	ACTCATAGAAGGGGAAGACCGTTCTTCTGGGTTTATAACTGATAAACTCAACAGAGAAGC	18474
Qy	955	cgagaagaagaagaagattgagttgcttggg-----gaatt	993
Db	18473	GAAATATGAAGGAGAAGAAGACAGAGATTGGAAGATAGCTGGTTTTAGACACAGCT	18414
Qy	994	gaaaaaaatggggaaaaatgatttcgtgctgcgcagttgaggttcttgcgccactgc	1053

Db	18413	TGAAGAGCTTGGGATGATTTGCTCTGCTGTGTTCCGAGATAGAGGTTTTGAGACACCGAGC	18354
Qy	1054	gttggagatgttttcgtagcagcatgttggttgaaactcggctgtgagagctgtgagttgcgg	1113
Db	18353	CATAGGTTGTTTTTGGACTCAATTTGGGTGGAGCTCATCTCGAGAGTTTGGTCTCGG	18294
Qy	1114	ggttcgggtgggtgggggtgcgcagctgggttgcacagcagcaaatgcgaagctgattga	1173
Db	18293	CGTTCCAGTGGTGGCGTTTCCGATGTGTGTCGATACGCCACCAAAATCGGAAGCTTTTGGG	18234
Qy	1174	ggatgcgtgggggacaggggtgagagtgagaaatgaagggcgtgggggttgatggatc	1233
Db	18233	AGAAATATGGAAGACACAGTGTGGGTGAGAGAGAACTCCGAAGCTTTAGTAGAGAGAGG	18174
Qy	1234	tgagatagagaggtgtgtggagatgtgtgatggatgggggtgagaagacgcaaatagtag	1293
Db	18173	AGAGATAAATCGCGTGTTTGGAAGCAGTGTATGGA-----GCCGAAATCGGTGAGCTGAG	18120
Qy	1294	agaaatgcataaaatgaagactttgcccagagaagccatgggagagatggatcttc	1353
Db	18119	GGAAACCGCAGAGAATGGAACGCTTTAGCAGCTGAAGCGGTGAGAGAAGGAGATCTTC	18050
Qy	1354	actcaagaatctcaagcctttct	1377
Db	18059	GGACAAGAAATGTGGAAGCTTTTGT	18036
RESULT	10		
LOCUS	ATCHRIV42/c		
DEFINITION	Arabidopsis thaliana DNA	DNA	PLN
ACCESSION	AL161542		16-MAR-2000
VERSION	AL161542.2	GI:7268303	
KEYWORDS	thale cress.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 195068)		
AUTHORS	EU Arabidopsis sequencing project.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-2000) MTPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemckem@mpi.biochem.mpg.de, mayer@mpi.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thaliana this fragment has an overlap with ATCHRIV41 at the 5' end and an overlap with ATCHRIV43 at the 3' end.		

	FEATURES	SOURCE
1.	100% Pure Cotton	USA
2.	Machine Washable	India
3.	Soft & Comfortable	Bangladesh
4.	Durable & Long Lasting	Pakistan
5.	Eco-Friendly	Vietnam
6.	Available in Multiple Colors	Thailand
7.	Lightweight & Breathable	Philippines
8.	Easy to Fold & Store	Malaysia
9.	Hypoallergenic	Singapore
10.	Perfect for Travel	Indonesia

source

exon

5

Qy 994 gaaaaaatggggaaaatagtttcgtgggtgcagttggaggtctcggcgcaacctgc 1053

```

gb:AW004332.1, A1997953.1, W43562, N96015, W43388"
/codon_start=1
/product="DEF (CLAL) protein"
/protein_id="CAB78598.1"
/db_xref="GI:7268304"
/translation="MASSAFAPSYIITKGGLTSDCKSTSLSSRSRLVTDLPSPCLK
PNNHSHNRRAKVCASAEKGEYYSNRPTPLDITINPIHMKNLVSEKELKQLSDELK
SDVIFNVKSTGHLGSSLGVELTVALHYIENTPQDKILDVGVCHQSPHKILIGRGK
MPTMRQTNGLSPTRGSESEHDCFTGHSSTIISAGLGMVGRDLKGNKNNVAVYIGD
GPTAGQAEAMNAGILSDSDIVILNDKNQVSLPTATLDGSPPPVGCALSALSRLQS
NPALRELVAEKGMTKQIGGPMHQAAKVDYARGMISGTGSSLSFEELGLYITGPVDG
HNTDQVAILKEVKTSTRTGPVLIHVTEKRGYPYAEARDDKYHGVYKFDPATGRQF
KTNKTQSTVTFYFAEALVAEAEVDKDVVAIIHAAMGGSTGLNLFQRRPPTCFDFVGAIE
QHAVTEAAGLCEGLKPCFCAIYSSFMQRAYQVYVHDVLDQLPVRFAMDRAGLVGADG
PTUCGAFVTFNACLPLNMIVMAPSDEADLFNNVATAVAIDDRPSCFRYPRTNGIGVAL
PPNKGVPTEIGKRTLKEGERVALIGYSAVQSCLAGAAMLEERGILNVAVADARECK
PLDRLIRSLAKSHEVLIITVEGSLGPGSHVQVFLALDGLLDGKLLKWRPWLPRDYI
DHGAPDQAEAGLMPHSIAATAALNIGAPREALF"
9917..12953
/gene="AT4g15560"
/intron
10067..10349
/gene="AT4g15560"
/number=1
10350..10475
/gene="AT4g15560"
/number=2
10476..10558
/gene="AT4g15560"
/number=2
10559..10874
/gene="AT4g15560"
/number=3
10875..10980
/gene="AT4g15560"
/number=3
10981..11252
/gene="AT4g15560"
/number=4
11253..11339
/gene="AT4g15560"
/number=4
11340..11613
/gene="AT4g15560"
/number=5
11614..11690
/gene="AT4g15560"
/number=5
11691..11998
/gene="AT4g15560"
/number=6
11999..12080
/gene="AT4g15560"
/number=6
12081..12362
/gene="AT4g15560"
/number=7
12363..12440
/gene="AT4g15560"
/number=7
12441..12725
/gene="AT4g15560"
/number=8
12726..12812
/gene="AT4g15560"
/number=8
12813..12953
/gene="AT4g15560"
/number=9
14413..14436
20338..24557
/gene="AT4g15570"
20338..20373
/gene="AT4g15570"
20374..20766
/gene="AT4g15570"
/number=1
20767..20852
/gene="AT4g15570"
/number=2
20853..20961
/gene="AT4g15570"
/number=2
20962..21012
/gene="AT4g15570"
/number=3
21013..21255
/gene="AT4g15570"
/number=3
21256..21367
/gene="AT4g15570"
/number=4
21368..21466
/gene="AT4g15570"
/number=4
21467..21572
/gene="AT4g15570"
/number=5
21573..21652
/gene="AT4g15570"
/number=5
21653..21750
/gene="AT4g15570"
/number=6
21751..21988
/gene="AT4g15570"
/number=6
21989..22117
/gene="AT4g15570"
/number=7
22118..22181
/gene="AT4g15570"
/number=7
22182..22274
/gene="AT4g15570"
/number=8
22275..22624
/gene="AT4g15570"
/number=8
22625..22678
/gene="AT4g15570"
/number=9
22679..22956
/gene="AT4g15570"

```

join(<20338..20373,20767..20852,20962..21012,21256..21367,
 21467..21572,21653..21750,21989..22117,22182..22274,
 22625..22678,22957..23015,23122..23206,23286..23765,
 24183..24269,24366..24557)
 /gene="AT4g15570"
 /note="similarity to SEN1 protein, yeast; Contains
 ATP/GTP-binding site motif A (P-loop) AAL3-20; contains
 EST gb:N63040"
 /codon_start=1
 /product="SEN1 like protein"
 /protein_id="CAB78599.1"
 /db_xref="GI:7268305"
 /translation="VGLSRKSFVLIQGPCTGKTOTILSILGAIHATPARVQSKGTD
 HVKRGQMTIOEKYNHGRASPMIIGVNPDAIMPEDGDDGFPPTSGNELKPEVNA
 SRKRVLVCAPSNALDEIVLRLSSGLRDENAOQTYTKIVRIGLKAHHSVASVSL
 DHLVAQRGSAIDKPKQGTGTGDIISRTALLEEAIIWGFNERLINVPATLSFGSAL
 LAKSNRGFDVVIDEAAQAVEPATILIPRCKQVFLVGDPNQLPRAIVSTVADQSGT
 TSNFERLQAGVPVRMLKATQYRMHPPELIPISQTPKPIVVSLLTTPMIASTYITQTIP
 LTENQIRSFPSKQFYEGALEDSDEIAQTRDMHKYRCGPGCFDHEGESHQPGA
 TGSVNLDEVEFYLLIYHRLVTMYPELKSQSLAIISPYNTQVKTFKDRKEMFGTEA
 EKVVVDINTVDGFQVGSAAATLAKSDPLWKNLIESAQNRNLFKVSKPLNNFFSEENLET
 MKLTEDMEIPDALYEDESLTPVAPYGGDDDFGDDGADQDDVAMAGED"
 20374..20766
 /gene="AT4g15570"
 /number=1
 20767..20852
 /gene="AT4g15570"
 /number=2
 20853..20961
 /gene="AT4g15570"
 /number=2
 20962..21012
 /gene="AT4g15570"
 /number=3
 21013..21255
 /gene="AT4g15570"
 /number=3
 21256..21367
 /gene="AT4g15570"
 /number=4
 21368..21466
 /gene="AT4g15570"
 /number=4
 21467..21572
 /gene="AT4g15570"
 /number=5
 21573..21652
 /gene="AT4g15570"
 /number=5
 21653..21750
 /gene="AT4g15570"
 /number=6
 21751..21988
 /gene="AT4g15570"
 /number=6
 21989..22117
 /gene="AT4g15570"
 /number=7
 22118..22181
 /gene="AT4g15570"
 /number=7
 22182..22274
 /gene="AT4g15570"
 /number=8
 22275..22624
 /gene="AT4g15570"
 /number=8
 22625..22678
 /gene="AT4g15570"
 /number=9
 22679..22956
 /gene="AT4g15570"


```

||||| ||| ||||||||| || ||| ||||||||| || ||| ||| ||| ||| ||| |||
Db 196262 TCCGCAATGATGATCAATGATGAACGCGAGCTTTTGAAGATGTTGGAAACAGG 196203

Qy 1192 ggtgagagt-----gagaatgaatgaaggggtgggtgatgatgatgatgatgat 1245
|| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196202 TGTAAAGAGTGATGCGAAGCAAGCAAGAGGAGTGTGTGTGGATAGTGAGGAGAT 196143

Qy 1246 gttgtggagatggtgatgatggtgggtgagaagagcaaaactagtgaagaaatccat 1305
|| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196142 ACGCCGTGCTATGAGGAAGTTATGAACACAGCAGCGGAGGATTTAGAGGAATCCAC 196083

Qy 1306 aaatggaagactttggccagagagccatgggagagatggtatcttcaatcaagaatct 1365
|| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196082 GAGGTGGAAGATTATACGGCGGAGGCTGTGAGAGAGGAGGCTTCTCTTAATCATCT 196023

Qy 1366 caagcctttcttcatca 1383
||| || ||| ||| ||| |||
Db 196022 CAAAGCTTTGTCGATGA 196005

RESULT 12
ATFCA4/c
LOCUS
DEFINITION ATFCA4 205065 bp DNA PLN 28-JUN-1999
Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
No. 4.
ACCESSION 297339
VERSION 297339.2 GI:5281015
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 205065)
Bevan,M., Stokema,W., Murphy,G., Wambutt,R., Pohl,T., Terry,N.,
Kreits,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
Mewes,H.W., Schueller,C. and Chalwatzis,N.
Unpublished
JOURNAL
2 (bases 1 to 205065)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schuellemips.biochem.mpg.de,mayeremips.biochem.mpg.de,project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
On Jun 30, 1999 this sequence version replaced gi:2244901.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with AtfCA3 at the 5' end and an
overlap with ATFCA5 at the 3' end.
FEATURES
source
1..205065
Location/Qualifiers
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
source
1..14029
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="cosmid YUP5B11.IB1G15/YUP5B11.IB1G19"
4257..7108
/gene="dl3766w"
4257..4662
/gene="dl3766w"
/number=1
join(<4257..4662,6078..7108)
/gene="dl3766w"

/note="strong similarity to fatty acid hydroperoxide lyase
- Capsicum annuum; contains EST gb:949688"
/codon_start=1
/product="hydroperoxide lyase (HPOL) like protein"
/protein_id="CAB45989.1"
/db_xref="GI:5281016"
/translation="NISAATITPLTAPPYNAGIVRLAVGWTIIGPFRQGPDKPEFT
RAEKYKSVFRNIPPTPPFGNPNPIVAVLDVKSFLHFDMDLDVKDRDVLIGDRP
SGFTGVRVGYLDTTEPKHAKRTIVALLLLRFAMETLKRSKVMQLKRNLAIF
WGTISEISNGAASYIFPLQRCIFSELCASLAGVADSPDIAENGWKTINTWLAIQ
VPTARGLVPOPLEEILLHTWPSLLIAGNYKKLYNFIDENAGDCLRIGQREFGLT
RDEALQNLLVILFGNAYGFSVFLPSLITGRITGNSGLQERIRTEVRRVCGSGDLNF
KTVNMEVLKSVVYETLAFSPVPLQFARAKDFQISSHDAFEVKKLLCCYQPLV
MDVANFDEPEEFKPDVVGVTGSELLNLYLWSNGPOTGTPSASNKQCAAKDIVTLTA
SLLVADLRLVDTITGDSGSIKAVVKAK"
4663..6077
/gene="dl3766w"
/number=1
6078..7108
/gene="dl3766w"
/number=2
8083..8712
/gene="dl3768w"
/number=1
8083..9807
/gene="dl3768w"
join(8083..8712,8790..8899,9225..9395,9521..9574,
9654..9807)
/gene="dl3768w"
/note="predicted"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB45990.1"
/db_xref="GI:5281017"
/translation="MSFFRSKTSKTETQPPSSYQQQQDAMYSYGVTVSDVLLQIHGC
RLFLNGSVAELAAGDFELVQVLDNSVALAMVVRIGNDLQWPVKDPPVKLDSRDY
LFTLPVKGDELYGITYFPDENDIVFVNSIELDDFLRNSCFSSPSSSSSSSS
SSSSVNGIDWKEFAPKIEDYNNVAKAIGAGTGHIIRGMFKCSNAYTNQVHKGGIM
ITKAEKNGASKKRATTNQINKLQRVRKLRATEKLSKTMNGCVGVSGVMGP
VKSIPGKAFFSMVPGEVLLASLDALSKETNSFCYVOSTRMYSERLIGESAGETKDV
LGTVGHAGTAWNVENIRKAFSPSSYSTGILKNASK"
8713..8789
/gene="dl3768w"
/number=1
8790..8899
/gene="dl3768w"
/number=2
8900..9224
/gene="dl3768w"
/number=2
9225..9395
/gene="dl3768w"
/number=3
9396..9520
/gene="dl3768w"
/number=3
9521..9574
/gene="dl3768w"
/number=4
9575..9653
/gene="dl3768w"
/number=4
9654..9807
/gene="dl3768w"
/number=5
complement(10065..10511)
/note="similarity to L.esculentum Tfm5 gene"
/codon_start=1
/product="glycine-rich protein like"
/protein_id="CAB10324.1"
/db_xref="GI:2244903"
/db_xref="SPTREMBL:O23398"
/translation="MKITKTRAHFVINYAAKLAVALLIITLLLNCTTVVAKGGGGAH"

```

exon	GGCGVHVSVGGGAHSAVGGGCHASGCGGCHASVGGGCHASGCGGCHAVGGGCHAGGGGGGHHCE EEGGHIGKGGGWHRRPATRNGSTLITLPVDDGSTTVTFMIIFFLFN" complement(10085..10511)
gene	/number=1 /genes="dl3770c" join(12306..12597,13233..13495,13554..13835,13944..14000, 15045..15290,15396..15802,16095..16280,16380..16453, 17015..17346) /genes="dl3775w"
CDS	/note="similar to hypothetical protein C02F5.7, Caenorhabditis elegans; contains EST gb:1217311;937779;315469" /codon_start=1 /product="hypothetical protein" /protein_id="CAB10325.1" /db_xref="GI:2244904" /db_xref="SPTREMBL:O23339" /translation="MDKPYGTASVSMGGIDRSAGKDIDILEMGVGEATLYPGLSYGENQ LRMGFTIRKYGILTSAGLLFTLLISAVVLPVNDLLTSGPGLILFLCIVPFIILLSE FSTYETKDYELHSRTSAIFYMSYLLTRFIAHYDIYWEFNFVVIWPLHYHQKHVPNLI LALFTYSGKDFVSGSCAMTEGCYLLPLNSPYCKEQLPETWNSKACKSCVFTNNLL FFLLRYLPGRIVQIALIILTSVWGLTAYTFMAAKKGDFSELGPIILFTSLILWVTS FIOMFFLPCTSVAVYGGFSALKRRKGRDSSPSSSRKRLTKDTHSGAENVSSSILT DTGLTALANGPRRIENLSLWCPNVSSVGLSLAQKCTSLKSLDQGCYVGDQGLAAV GKFKQLEPLNRFCEGILTGVDTLVVGGCKSLKSIGVAAAKITDLSLEAVGSHCK LLEVLYLDEYTHDGKLIATAQGCCHRLKNKLOCYSDTDAFAAAGELECTSLERLALY SPQHTDGMRAIAGKSLKDLTSLDSCYFVSCKGLEIAHAGCKELEKEVEINGCHNIG TRGIEAIGKSCPLRKLKALLYCQRIGNSALQETGKGLKAGTFDHFKNQIDMPLAE GEGCPMLKDVLISHCHHTDNLGNLHVQKCKLLETCHVYCPGCIITSAGVATVVSQCPH IKKVLTEKKWVTERTRRAGSVISYLCMDL"
exon	12306..12597 /genes="dl3775w" /number=1
gene	12306..17346 /genes="dl3775w"
intron	12598..13232 /genes="dl3775w" /number=1
exon	13233..13495 /genes="dl3775w" /number=2
intron	13496..13553 /genes="dl3775w" /number=2
exon	13554..13835 /genes="dl3775w" /number=3
intron	13836..13943 /genes="dl3775w" /number=3
exon	13944..14000 /genes="dl3775w" /number=4
intron	14001..15044 /genes="dl3775w" /number=4
source	14700..47605 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="cosmid YUP5B11.IB11L2/YUP5B11.IB1A8" 15045..15290 /genes="dl3775w" /number=5
intron	15291..15395 /genes="dl3775w" /number=5
exon	15396..15802 /genes="dl3775w" /number=6
intron	15803..16094 /genes="dl3775w" /number=6

exon 16095..16280
 /gene=U137/5w*
 /number=7

Query Match	12.2%	Score 184.4	DB 7	Length 205065
Best Local Similarity	60.0%	Pred. No. 4.2e-30		
Matches 371	Conservative 0	Mismatches 226	Indels 21	Gaps 3
QY 787	taactgcgtgagtggtttgacacgaagccgaatacttcctggtggtggtatgctgctgttgg	846		
Db 47149	TGAATACATAGAGTGGTGTGGATCAATAAGCGGAATCGCTGCTGCTTTATGTTTGGTCGCGG	47090		
QY 847	gagcgcttttgaggttttccaaaggccacaaattggaagagattgggaagggctcattagcctg	906		
Db 47089	GACGCTTCGGCTGTGACRAGAACAAGCTTGTGGAGCTTTCTAAGCGTTCATACAAAG	47030		
QY 907	cgaagggccgtttttatggatgatacagaagaag-----agaatgacgcagcgcga	957		
Db 47029	TCGGAGACCATCTCTGTGGGTGATTCAGGATAAGTCGTACAGAAATAAAGAGATGAGCA	46970		
QY 958	agaaagaagaagaagtg-----tgagttgcattgggaattgaaaaaaatggggaaaaat	1011		
Db 46969	AGAGAAGGAAGAAGATTGCATAAGTACTTTTCAGACAAGAGCTCGATGAGATAGGAATGGT	46910		
QY 1012	agtttcggtggtctgcagttggaagttctggcgacacctgcgttgggatgtttcgtgac	1071		
Db 46909	GGTTTCATGGTGTGATCAGTTTACGGTTTGAATCATAGATCGATAGGTTGTTTCGTGAC	46850		
QY 1072	gcatttggttggaactgcgtctggagagcttgagtttcgggggttcctgggtgggtgaggt	1131		
Db 46849	GCATTCGCGGTGGAACCTCAGCTGGAGAGCTTGGTTTCAGGAGTTCGGGTGGCGGTT	46790		
QY 1132	gcgcagtggtttgatcatcacagcagaatgcgaagctgattgaggatgcgtgggggacagg	1191		
Db 46789	TCCGCAATGGAATGATCATGATGATGAACGCGAAGCTTTTAGAAGGATTTTGGAAAAACGG	46730		
QY 1192	ggtgagagt-----gagaatgaaatgaaggggtggtggtgattgagctctgagatagagag	1245		
Db 46729	TGTAAGAGTGTGGAGAGAAGGAAGAGAGTGTGTGGTGGATAGTAGGAGAT	46670		
QY 1246	ggtgtgagatggtgatgagtgagtggtgagagaagcaactagtagagagaagaatgccat	1305		
Db 46669	ACGGCGGTGCATTGAGGAAGTTATGGAAGACAAGCGGAGGAGTTTACAGGAATGGCAC	46610		
QY 1306	aaattggaagactttggccagagaagccattggagagaggtgagcttcactcaagaatctt	1365		
Db 46609	GAGGTGGAAGGATTTAGCGCGGAGGCTGTGAGAGAAGAGGAGCTCTCTCTTAATCATCT	46550		
QY 1366	caacgccctttcattca 1383			
Db 46549	CAAAGCTTTGTGCGATGA 46532			

RESULT 13	1589 bp	PLN	02-JUN-1997
ATU01293	1589 bp	PLN	02-JUN-1997
LOCUS	Arabidopsis thaliana	UDP-glucose:indole-3-acetate	
DEFINITION	beta-D-glucosyltransferase (laglu) mRNA, complete cds.		
ACCESSION	U01293		
VERSION	U01293.1	GI:2149126	
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Arabidopsis thaliana		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
TITLE	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core		
	eudicotis; Rosidae; eurosids II; Brassicales; Brassicaceae;		
	Arabidopsis.		
	1 (bases 1 to 1589)		
	Graham, R.A. and Thornburg, R.W.		
	Data Sequence of UDP Glucose:indole-3-acetate		
	Beta-D-Glucosyltransferase from Arabidopsis thaliana (Accession No.		
	U01293) (PGR97-044)		
JOURNAL	Plant Physiol. 113, 1004 (1997)		

2 (bases 1 to 1589)
 Thornburg,R.W. and Graham,R.A.
 Direct Submission
 Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State
 University, Ames, IA 50011, USA
 Location/Qualifiers
 1..1589
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 26..1589
 /gene="laglu"
 26..1450
 /gene="laglu"
 /EC_number="2.4.1.121"
 /function="transfers the glucose moiety from UDP-glucose
 to indole-3-acetic acid to form ester linked IAA-Glc"
 /note="IAA-glucose synthase"
 /codon_start=1
 /product="UDP-glucose:indole-3-acetate
 beta-D-glucosyltransferase"
 /protein_id="AAB58497.1"
 /db_xref="GI:2149127"
 /translation="MANNNSSTGPHFLFVTFPAQGHINPSLEAKRLAGTIGSARY
 TFAASIASTENRFEFTSENPETLFIATYSDGHDDGFKSSAYSDKSRODATGNFMSER
 RRGTEUTLEIENRKNRPFTCVVYITLLTWAELEHFLPSALLWQVPTVFISF
 YHYNGKDIASEMANTPSSIKLPSPLLTVTRDIPSFIVSNVYAFLLPFAREQIDS
 LKEINPKILINTFOELEPEAMSVPDNFKIPVGPPLTLATDFSSRGEYIEMLDQIS
 DSSVYSFGTAVLSKQOLVELCKALIQSRPPLWITDKSYRNKDEQEKEDCIS
 SSEKSFGEIGVYSWDCDQFVLNHRSIGCFVTCGWNSTLESLSVGPVVAFPQWDO
 MTNAKLEDCKKTVGRVMEKKEEGVVVDSEIRCIIEVMEKAEFRGNATRWKD
 LAAGAREGSGSSNHLKAFVDEHM"
 1568..1575
 /gene="laglu"
 1588..1589
 /gene="laglu"
 /note="24 A nucleotides"
 BASE COUNT 433 a 342 c 375 g 439 t
 ORIGIN

 Query Match 12.0%; Score 181.2; DB 8; Length 1589;
 Best Local Similarity 59.7%; Pred. No. 1.7e-29;
 Matches 369; Conservative 0; Mismatches 228; Indels 21; Gaps 3;

Qy	787	taactcgtggaagtgttgacacagacgaacgaatttcggtgtgtatgtctcttgg	846
Db	823	TGAATACATAGAGTGGTGGATCTAAGCGGATTCGCTGCTTATGTTCGTGG	882
Qy	847	gagcgttttgaggtttccaaaggccaaatggagagattgggaaggcgtattagcgtg	906
Db	883	GAGCGTTGCGGTGTGAGCAAGAACACAGCTTGTGAAGCGCTTGATACAAAG	942
Qy	907	cgaaggcgtttttatgatgatacagagacag-----aagaatgacagcgga	957
Db	943	TCGGAGACCATCTCTGTGGGTGATTACGGATAAGTCGTACAGAAATAAAGAGATGACCA	1002
Qy	958	agaagaagaagagagttagttgcattgggga-----attgaaaaaatgggaaaaat	1011
Db	1003	AGAGAAGGAAGAAGATTGCATAAGTAGTTACAGAGAAGAGCTTCGATGAGATAGGAATG	1062
Qy	1012	agtttggtggtcctgcaggttggagggttcctgcgcacccctgcgtgggagtgttcgtgac	1071
Db	1063	GGTTTCATGGTGTGATCAATTTAGGTTTGAATFATAGATCATAGTATGTTTCGTGAC	1122
Qy	1072	gcattgtaggtggaactcggctgtggagagcttgagtgcggggtccggtggtggtg	1131
Db	1123	GCATTCGGGTGGAACCTACCTACCTGGAGAGCTTGGTTTCAGGAGTTCCGGTGTGGCGTT	1182
Qy	1132	gcgcgagtggtttgatcagacagcgaatcgcaagctgattgaggtatcgctgggggacagg	1191
Db	1183	TCGCAATGGAATGATCAGATGACGAACGCCAGCTTTTGAAGATTTTGGAAACAGG	1242

numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F623, F536, and T1316, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

```

FEATURES
    source            Address all correspondence to: atetigr.org.
    1..99188          Location/Qualifiers
    /organism="Arabidopsis thaliana"
    /cultivar="Columbia"
    /db_xref="taxon:3702"
    /chromosome="II"
    misc_feature      <1..1802
    /note="Sequence from clone T9H9"
    mRNA              complement(<1064..>2005)
    /gene="At2g31720"
    gene              complement(<1064..>2005)
    /note="At2g31720"
    CDS               /note="T9H9.24; predicted by genscan"
    complement(1064..2005)
    /gene="At2g31720"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="A032277.2"
    /db_xref="GI:5598601"
    /translation="MYRGDQLNGGKRALDILCVATERTVSAPAYEEIGETSTBG
    KRPOKVVNRKRLNRPVIVTDEDTPTMLIDVNRVNGMDAKLIFVKVLPSNDVDE
    LQTRLMNPKQILDMDFLANEELEKIDRHKKISASDKGADVIVNKSGLQRLKLR
    WDMTSTYVLGDKWNVVTDILQRTLRISFSPDMLFFAFVLSGDPDPAPTECL
    NLLAKLEETTCLEAQEARNRSSIVSDTELDLELR"
    misc_feature      /note="Sequence from clone F20M17"
    mRNA              complement(1803..>99188)
    /gene="At2g31730"
    gene              complement(1803..>99188)
    /note="Sequence from clone F20M17"
    complement(3588..4265)
    /gene="At2g31730"
    CDS               complement(3813..3973,4058..4094,4189..4260))
    /note="unknown protein"
    /codon_start=1
    /protein_id="A032299.1"
    /db_xref="GI:4887763"
    /translation="MQGDTASVLLGSMQYIQFQVRCSLSNPFTYRVEEYSLRS
    KGLCLVPLEYTSVADTNGADIWAPVKPTTSSHAFNLSSNSPQ"
    mRNA              complement(join(<7083..7286,7407..7557,7647..7813,
    7921..7995,8070..8150,8243..8367,8449..8698,9242..9289,
    9386..9469,9576..9833,9973..10145,10486..10555,
    10653..>11039))
    /gene="At2g31740"
    gene              complement(<7083..>11039)
    /note="predicted by genscan"
    CDS               complement(join(7083..7286,7407..7557,7647..7813,
    7921..7995,8070..8150,8243..8367,8449..8698,9242..9289,
    9386..9469,9576..9833,9973..10145,10486..10555,
    10653..11039))
    /gene="At2g31740"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="A032298.1"
    /db_xref="GI:4897762"
    /translation="MCKKKGKNAASATDDFLQTLDEFTSKENWDKFFTLRGNDSPFM

```

```

YAEPQQLHDSLLPLLDSSSSSSSLQILYPCGNSRLTEHLYDAGFRDITNVDPSKV
VISMLRNIRKPELRNRMVMDITKMLADESFDTVLQDGLADLMDEPEVALLFSRER
FGKMNVHSIAQRKSLKTYMVAEKENSVLHEITSAPFVLSGRNDSQSGMREAV
ESENKIRDCNNGSLLYSHEDLKLIGDGLAELTEGRRIKFTLGGQGSNFSYRAVLL
DAQKOTPFYVYCGVLPKTRAHMLFCSEBQWQVYVESQAARLIMVFLDSSHGA
TMDEGIIYSPMTKVVVDVYVESAPNSLEDLSTSDLAFLRLVFKRTGLIQSEALL
VEDGIIIEOSQKBTKNVSKQKNGKNQONPSRPLMRVSHDVLYASSYHTGIIISGFT
LVSSYLKASCSCTMYKTVVIGLAGLLPMLHGCLPFESIEAVELOPMLSVCKDYF
GFTONRLKVIADGINKFIRDIINSEASSESNIGSNGSDTTHNTQGGICPCDLIID
VDSADSGGLTSPASDFIETILLSVKQALPOHGLFIVNLVTRSOVKDMVSRMKKV
FDHLGLQLEEDDDVNVVFLGLCSVTSENDIPESAVILEGLLKCQRLETKQSIIDA
TKKLKWK"
repeat_region       10152..10235
/rpt_family="(TAA)n"
complement(12661..12688)
repeat_region
/rpt_family="(TAAAA)n"
mRNA               join(13281..13986,15175..15987)
/gene="At2g31750"
13281..15987
gene              /note="At2g31750"
/gene="At2g31750"
CDS               join(13342..13986,15175..15900)
/gene="At2g31750"
/codon_start=1
/product="putative glucosyltransferase"
/protein_id="A032297.1"
/db_xref="GI:4887761"
/translation="MGEKAKNVLFSPFIQGHINPLQLQFKSKRLSKNVNVTFLTTS
THNSILRRAITGGATALPLSFVPIIDGFEEDHPSTDTSPDYFAKFQNVSRSLSELS
SMKPNNAVYDCLPVLQVCKRHGPAASFFQSSSTVNATYIHFGDFEKVEFQND
VVLPAKPLKGNLDPVFLVDNNLCPLPELISSQFVNVDDIDFLVNSDFDEVEVLQ
WMKNQVPKNTGPMTPSYLQKCLAGDKDYGINLFAQVNECLDMLDSKPPGSVYVS
FGSLAVLKDDOMLEVAAGLKQCHNPLHVNRETEKLLPSVIEDICDKGLIVNMSQ
LQVLAHISGCFMTHCWNSTLEALSGLVALGMPAYSDDQPTNAKFIEDVWKKVGRVK
ADQNGFVPEKEIVRCVGEVNMEDNSEKGEIRKRNLMFEAREALSDGNSDKNIDEF
VAKIVR"
repeat_region       22837..22997
/rpt_family="(CAAAA)n"
repeat_region      23089..23130
/gene="At2g31760"
mRNA               join(<24523..25698,25968..26061,26146..26420)
/gene="At2g31760"
gene              <24523..26420
/gene="At2g31760"
CDS               join(24523..25698,25968..26061,26146..26420)
/gene="At2g31760"
/codon_start=1
/product="putative ARI-like RING zinc finger protein"
/protein_id="A032296.1"
/db_xref="GI:4887760"
/translation="MDYSDDDMDINSGEENSDGGNESYNNVNAVDTIILSEKSYV
IIEKEDILKORDIERVSTIFLSQVEAIVLLHVMCVSKLEDFTEERTKTV
GILKEPVYDNGTEVDIQCCIFCESTRKEIASVSGHPYKCTWGTITTKIEDGPG
CLRVKCPSPCYAVVGQMDIDEVTEKKDKDYVFLRSYVEDGKKMKWCPSPCECA
VFCGSSGVDAVLCSYRFGNCSEDAHPVDCETVSKVIFKNODESNKNWILANS
PCPKAKRPIKSHGCCNMTCSACGHRFCWICGKSYSDHVCANNVDADHKRFTLLQ
SEIKRTHYVVRWENSGSLKAMSDLEKFSQVLKQLSDNCKPKIDQLQFIDAWLQ
LIECRVRLKWTYATGYTIDNLAKRPLFETLQGEAETGRHLCRAENLAKOFFRSED
SEGCF"
repeat_region       27052..27180
/rpt_family="(TAAAA)n"
mRNA               join(<27609..28814,29071..29164,29249..29580)
/gene="At2g31770"
<27609..29580
gene              /note="At2g31770"
/gene="At2g31770"
CDS               join(27609..28814,29071..29164,29249..29580)
/gene="At2g31770"
/codon_start=1
/product="putative ARI-like RING zinc finger protein"
/protein_id="A032295.1"

```


FEATURES

```

Db 823 AAGAGGATAAAGGCTACGGCTCGAGCTGAT---GAAGCCGCTCTCGAGGAGGTATGG 879
QY 798 agtgggttgacacgaagccgaatacttcggtggtgtagtgcgtttggtggagcgctttga 857
Db 880 AGTGGTTAGACACTAAGCTGAGCAAGTCGGTGGTTTTTTCGTTTGGTTCCCTTTGGGA 939
QY 858 ggtttccaaaggccaaatggaagagattgggaaggctattagctgcggaaggccgt 917
Db 940 TCCCTCTTGGAGAACCACTCGCTGAGGTGGCAAGGCGTTACAAGAAATCCAACTTTAACT 999
QY 918 ttatatgatgatacgagacacagaagaatgacgacgagcggaagaagaagaagagttga 977
Db 1000 TCTTGTGGGTGATCAAAAGAGCTCATATAGCGAAATACCAAG----- 1044
QY 978 gttgcatggggaattgaaaaaatgggaaaaatagtttcggtgctgcagttggagg 1037
Db 1045 GGTTTGTGGAAGCTTACCAAGACAGAGCGTGTGCTTTCTTGGTGTAAACAGCTTTGAGG 1104
QY 1038 ttcggcgcacctcgcttgggatgttcgtagcgcattgtggtgggaactcggctgtgg 1097
Db 1105 TTTTAGCTCATGTATCGATAGGTTCGTTTGTGACTCACCTGCGGTTGGAACTCGACGTTGG 1164
QY 1098 agagcttgagttcggggttcggtggtggtggtggtggtggtggtggtggtggtggtggt 1157
Db 1165 AAGGATTGACTTTGGGAGTTCCGATGCTTGGTGTGCTCAGTGGAGTGATCAGATGAATG 1224
QY 1158 atgcgaagctgattgaggtatcgctggtgggacaggggtgagagtgaagaatgaatgaagg 1217
Db 1225 ATGCTAAGTTTGTGGAGAGGTTTGGAGAGTTTGGGTATAGAGCGAAGGAGGAAAGCTGGGG 1284
QY 1218 gtggggttgatggatctgagatagagagggtggtggtggtggtggtggtggtggtggtggt 1277
Db 1285 GAGGAGTTGTGAAGAGCGATGAGGTGCTCAGGTCTTTGAGAGGAGTGATGGAAGGAGAGA 1344
QY 1278 agagcaactagttagagaaaaatgccataaaatgggaagactttggccagagaagccatgg 1337
Db 1345 GTAAGTGTGGAGATTAGAGAGAGTTCTAAGAAGTGAAGAGATTGGCTGTGAAGCGCATGA 1404
QY 1338 gagaggatggatcttcactcaagaatctcaagcccttct 1377
Db 1405 GTGAAGGAGGAAGCTCTGATCGGAGCATTAATGAGTTTGT 1444

```

Search completed: August 1, 2000, 12:39:50
Job time: 54252 sec


```
QY 301 aaactcctcttcaacaacacacagctcaggttcgtctactcttccaccctctttgcatg 360
DB 301 AAACCTCTCTTCAACAAACACGACGTCAGTTCGTCGCTACTCCACCTCTTTGCAATG 360
QY 361 ggcgcgaggtggtcgctgagtcctccaggtcccgagcgccctctctctggtcgagccgcg 420
DB 361 GCGCGGAGGTGCGCGTCCAGTCCCGAGTCCCGAGCGCCCTCTCTCTGTCGAGCCCGC 420
QY 421 caccgtgctgcatatatacttaacttaactcaacggctacgcagacgagatcgagccgg 480
DB 421 CACCGTCTGTGCATATATTTACTTTCCTACTTCAACGGCTACGCAGACGAGATCGAGCGCG 480
QY 481 ttccgacgaattcagctccctcggtctccaccctggagcgagcagctctccgacctt 540
DB 481 TTCCGACGAATTCAGCTCCCTCGGCTTCACCCCTTGGAGCAGCGCACTCTTCGACCTT 540
QY 541 tctctcccgagagacacccgagagatcccggttgatgatgaaggagaagcttggaaacttt 600
DB 541 TCTCTGCCGAGACACCGGAGAGATTCGGGTGATGATGAAGAGGAAGCTTGGAAACTTT 600
QY 601 agacggtgaagagagcgaagtgttggtagaacgcttgatgctgttggagcccgatgc 660
DB 601 AGACGGTGAAGAGAGCGGAAGTGTGTGTGAACAGCTTTGATGCTGTGGAGCCGATGC 660
QY 661 actcaaggctatgatggtatgattgattcggtatcggtatcggttgatctccctcgacct 720
DB 661 ACTCAGCGCTATTGATAGGTATGAGTGTGATGATCGGGATCGGGCGGTGATTCCTCCGCTT 720
QY 721 cttggacggcgagatccctccgaacgctcttacgcgcgcatcttttcgaaaaatcgga 780
DB 721 CTTGGACGGCGAGATCCCTCCGAACGCTTACGSCGCGATCTTTTCGAAAAATCGGA 780
QY 781 ggaataaactcgttgagtggttgacacgaagccgaactctcgttggtgtagtgctc 840
DB 781 GGAATAAATCTGCTGGAGTGTGTGACACGAGCCGAATCTTCGGTGGTGTATGTGTC 840
QY 841 gtttggagcgttttgaggtttccaaagccacaaatggaagagattgggaagggtctatt 900
DB 841 GTTTGGAGCGCTTTTGAGGTTTCCAAAGGCACAAATGGAAGAGATGGGAAGGCGCTATT 900
QY 901 agcttcggaagccgcttttatgtagatcagaaacgaagaatgacagcgcgaaga 960
DB 901 AGCTTCGGAAGCGCGTTTTTNTGATGATACAGACAGAAAGATACGACGCGGAAGA 960
QY 961 agaagaagaagttgagttgcttgggaattgaaaaaattggggaaaaatagtttcgtg 1020
DB 961 AGAAGAAGAAGATTGAGTTGCATTTGGGAATTTGAAAAAATGGGCAAAATAGTTTCGTG 1020
QY 1021 gtgctcaggttgaggttctggcgaccctcggttgagatgtttcgtgacgcattatgg 1080
DB 1021 GTGCTCCAGTTGAGGTCTCTGCGCACCCCTCGGTTGGATGTTTCGTGACGCATTGTGG 1080
QY 1081 gttgaaactcggctgtggagagcttgagttgcggggttcgggttgggtgcgcgagtg 1140
DB 1081 GTGAACTCGGCTGTGGAGAGCTTGAGTTGCGGGGTTCGGGTGGTGGCGGTGCGCAGTG 1140
QY 1141 gtttgatcaagacacgatcgaagctgattagatcggtgagggacaggggtgaagat 1200
DB 1141 GTTTGATCAGACGACGAATCCGAAGCTGATTGAGGATGCTGTGGGGGACAGGGGTGAGAGT 1200
QY 1201 gagaatgaatgaagggttgggttbatggtatgatagatagaggtgtgtggagatggt 1260
DB 1201 GAGAATGAATGAAGGGGTGGGTTGATGGATCTGAGATAGAGAGGTGTGTGGAGATGGT 1260
QY 1261 gatgattgggggtgagaagacaaactagtagagaaaaatgccataaataatggaagacttt 1320
DB 1261 GATGATGGGGGTGAGAGAACAACTGGTGAGAGAAATGCCATAAATAATGGAAGACTTT 1320
QY 1321 ggcagagaagccatgggagagatggattcttcaatcaagaatctcaacgccttcttcca 1380
DB 1321 GGCCAGAGAAGCCATGGAGAGGATGGATCTTCACTCAAGAAATCTCAACGCGCTTCTTCA 1380
```

```
QY 1381 tcaagttgcaggtgcttaataacacaaaaatgggtttcccaacttttaactactcaaacacg 1440
DB 1381 TCAAGTTGCAGGTGCTTAAATACACAAAATGGCTTCCCACTTTAATCTACTCAAAACCG 1440
QY 1441 gtcaaaaataataccccctccactctcttcttatctcaactatcacattataaatttagt 1500
DB 1441 GTTCAAAATAATATCCCTTCCACTTCTTCTATTTCACATATCACATTTAATAATTAGT 1500
QY 1501 aacaaa 1506
DB 1501 AACAAA 1506

RESULT 2
X02827
ID X02827 standard; cDNA; 1474 BP.
AC X02827;
DE 14-MAY-1999 (first entry)
DT WO905287 Seq ID 2.
KW Plant; flavonoid 5-transglycosylation activity; 5tg; variety; colour; ds.
OS Perilla frutescens.
FH Key Location/Qualifiers
FT 29..1360
FT CDS /*tag- a
FT /product= "protein with flavonoid 5-transglycosylation
FT activity"
PN WO905287-A1.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR.) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI: 99-142940/12.
DR P-FSDB: W92948.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Disclosure; Page 53-56; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
CC Sequence 1474 BP; 369 A; 348 C; 438 G; 319 T;
SQ

Query Match 88.2%; Score 1329.4; DB 1; Length 1474;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 1403; Conservative 0; Mismatches 36; Indels 32; Gaps 2;

QY 2 aaatttccacaaaaatgggtccgcgcgcgtgctgtagaagcgtttcctgcgcaagcc 61
DB 14 AAAATTTCCACAAAATGGTCCGCGCGCGTGTCTAGCAACGTTTTCGCGCGCAAGGC 73
QY 62 cacataatcccgccctccaattcgcgaagagactcctaaagcggcgactgaagtcaca 121
DB 74 CACATAAATCCGCGCCCTCCAATTCGCCAAGAGACTCTAAAGCGGCACGTGACGTCAGC 133
QY 122 tttttcacgagcgtttatgatggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 181
DB 134 TTTTTCACGAGCGTTTATGATGCGCGCGCGCATGGCCAAACACAGCCTCCGCCGTCGCGGA 193
QY 182 aaccacccggccctcgacttcgtggcggttctccgacggctacgacgagcggtcgaaagcc 241
DB 194 AACCCACCGCGGCTTCGACTTCTGTTGGGTTCTCCGAGCGGTACGACGAGCGGTGAAGGCC 253
QY 242 tgcggcagcgggaagcgcctacatgtccgagatgaagccgcgcgcgcgcgcgcgcgcgcgc 301
DB 254 GCGCGCAGCGGGAAGCGCTACATGTCCGAGATGAAGCGCGCGGTCCGAGCGCTTAAGA 313
QY 302 aacctctcttcaaacaccacgacgtcacgttcgttcgtctactctccacctcttctgcatgg 361
DB 314 AACCTCTCTTCTCAACACGACGACGTCACCTTTTCGTGCTCTACTCCACCTCTTTTCATGG 373
```

```

Qy 362 gcggcgaggtgagcggtgagtcacaggtcccgagagccctctctctggtcgagccgccc 421
Db 374 GCggcgaggtgagcggtgagtcacaggtcccgagagccctctctctggtcgagccgccc 433
Qy 422 accgtgtgtgcatattactcttacttcaacggctacgcagacgagatcgagcgcggt 481
Db 434 ACCgtgtgtgcatattactcttacttcaacggctacgcagacgagatcgagcgcggt 493
Qy 482 tccgacgaattcagctccctcctgggttccacccctgagcagcgagcttccgaccttt 541
Db 494 TCCAAATGAATTCAGTCCCTCGGCTTCATCCCTGGAGCAGCGCATCTTCGACGCTTT 553
Qy 542 ctgtctccggagacacccggagagatccggttgatgatgaggaagctggaaacttta 601
Db 554 CTGCTCCCTGCGAGCGCGGAGAGATTCGGTTGATGATGAGGAGAGCTGGAACTTTA 613
Qy 602 gacggtgaagagaagcgaaagctgtgtgtaacacgctttagtcggttggagcccgatgca 661
Db 614 GACGGTGAAGAGAGCGGAAGATATTGTGTGAACACAGCTTGTATGCGTTGGAGCCGATGCA 673
Qy 662 ctacggctattgatlaggatgattgattcggtggtatcgggcgtttgattccctcgccttc 721
Db 674 CTCACGGCTATTGATAGGTATGAGTTGATCGGGATCGGGCGTGTGATTCCCTCCGCGCTTC 733
Qy 722 ttggacggcgagagatccctccgaacgctctacggcgcgctcttttcgaaaaatcgagg 781
Db 734 TTGGACGGCGAGAGATCCCTCCGAACGCTTTACGGCGCGCATCTTTTCGAAAAATCGGAG 793
Qy 782 gagaaataactcgtggtggtgtgacacgaagcgaacatcttcggtggtgattgtctg 841
Db 794 GAGAATAACTGCTGGAGTGTGATGTAACTCGAAGCGGAATCTTCGGTGGTGTATGTGTCG 853
Qy 842 ttgggagcggttttgaaggtttccaaagacacaaatggaagagattgggaagagctattta 901
Db 854 TTTGGAGAGCGTTTGAAGTTTCCAAAGGACAAATGGAAGAGATTGGGAAGGGCTATTTA 913
Qy 902 gctcggaagcggttttttattggtatgatalacgagaacgagaagatcgagcgcc----- 955
Db 914 GCCTCGGAAGAGCGCCCTTTTATGATGATATACGAGAACAGAAATGACGACGCGGAAGAA 973
Qy 956 gaagaagaagaagaagattgattgattggaattgaaataatgaaataatgaaataatgatt 1015
Db 974 GAAGAAGAAGAAGAAGTGTGATGTTGCAATGGGAATTTGAAAAAATGGGGAAAAATGATG 1033
Qy 1016 tcgtgtgtctcgagttgaggttctgtgacacccctggttgggatgtttcgtgacgcat 1075
Db 1034 TCGTGTGTCTCGCAGTTGAGGTTCTGCGCACCTGCTGTTGGGATGTTTCGTGACGCAT 1093
Qy 1076 tctggtgaaactcggctgtgagagcttgattgctgggttccggtgtggtgagcggtgccc 1135
Db 1094 TGTGGGTGGAACCTCGGCTGTGAGAGCTTGTAGTTGCGGGATTCGGGTGGTGGCGGTGCGG 1153
Qy 1136 cagtgtgttgcagacacgaatgcgaagctgattgaggtgctggtggggagacaggggtg 1195
Db 1154 CAGTGTGTTGATCAGACGACGAATGCGAAGCTGATTGAGGATGCTGGGGGACAGGGGTG 1213
Qy 1196 agagtgaagaatgaatgaaggggtggtggttgcgtgagatgagatagagaggtgtgtggag 1255
Db 1214 AGAGTGAGAAATGAATGAAGGGGTGGGGTGTGATGGATGTGAGATGAGAAAGGTGTGGAG 1273
Qy 1256 atggtgatgagtggtggtgagagacaaactagtgagagaaatgcatataaatggaag 1315
Db 1274 ATGTTGATGATGGGGGTGACAAAGACCAAACTAGTGAGAGAAATGCCATCAATGGNAG 1333
Qy 1316 actttggcagaagaagcaatggagagagatgattcttactcaagaatctcaacgctttt 1375
Db 1334 ACTTTGGCAGACAAAGCCATGGGATGGATGGATCTTCACTCAAAATCTCAACGGCTTT 1393
Qy 1376 ctctcaaatgttcagctgctctaaacacaaatggctttctccacttttaactactcaaa 1435
Db 1394 CTTGCTCAAGTTGCA-----CACTTTTATCTGCTCAAA 1472

```

```

Qy 1436 caccgttcaataataatataatcccttccactt 1466
Db 1428 CAGCGGTTCAATAATAATATATCCCTTCCACTT 1458

RESULT 3
ID X02828 standard; cDNA; 2062 BP.
AC X02828;
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 3.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Verbena hybrida.
FH Key Location/Qualifiers
FT 25..1411
FT CDS
FT /product= "protein with flavonoid 5-transglycosylation activity"
FT W0905287-A1.
PN 04-FEB-1999.
PD 16-JUL-1998; J03199.
PF 25-JUL-1997; JP-200571.
PR (SUNR ) SUNTORY LTD.
PA Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M.
PI WPI; 99-142940/12.
DR P-PSDB; W92949.
DT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT used to transform plants for improvement of plant coloration
PS Disclosure; Page 56-60; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 2062 BP; 608 A; 354 C; 508 G; 592 T;

```

```

Query Match 46.5%; Score 700.2; DB 1; Length 2062;
Best Local Similarity 71.3%; Pred. No. 3e-188;
Matches 1029; Conservative 363; Indels 51; Gaps 6;

Qy 2 aaaaattcccaaaaaatggtccgcccgcgtgctgctgtagcaacggtttctctgcaagggc 61
Db 11 AAAAAATAAAAAAATGAGCAGAGCTACGTCCTCTTGGCCACATTTCCCGACACAGGGA 70
Qy 62 cacataaatcccgccctcccaattcgccaagaagactcctaaagcggcgactgacgtcacca 121
Db 71 CACATAAATCCCGCCCTTCATATCGCCAAGCGCTCTCGCAATGCCGACATTCAGATCACA 130
Qy 122 ttttccagagctttatgcatggtcgccgcgtatggtcccaacacagctccgcgctgcgga 181
Db 131 TTTCTTACCAGCGCTTACGCATGGCGCGCATGTCCAGAACCAGCGCTGGGTCAAAACGG- 189
Qy 182 aacccacgggctcgacttcgtggttctccgacggtctacgacgacggtgtaagccc 241
Db 190 -----GCTCATCAATTTTGTGCTTTCCGACGGGTATGACGAGGGTTACAGCCC 241
Qy 242 tgcggcgaagcgaagcgtacatgtcccgagatgaaagcccgcggtccgagggccttaag- 300
Db 242 GGAGACGATGGGAAGAACACTACATGTGGGAGATGAAAAGCAGAGGTATAAAGCGCTTGAGC 301
Qy 301 -----aaacctctctcaaacaccacgactcagctcgctctac 343
Db 302 GATACATCTTCACGCCAATATGTGCATCAAAAACACCAAAATCACTGTCGTGGGTAC 361
Qy 344 tccaccttttgcattggcggtgaggtggtggtggtggtggtggtggtggtggtggtggt 403
Db 362 TCCACATCTTTTCATGGGGCGGCAAGGTGGCGGTGAGTTCCATCTCCGAGGCGCTA 421
Qy 404 ctctggtgagcccgcccgctgctgcatatattacttcttctcaacgctacgca 463
Db 422 CTCGGATTGAGCCAGCTAGGGTGTGGATATTTTACTTTTATTTTCAACGGCTATAGC 481

```


QY	624	tgttggtgaacacggtttgtagtgcttgaggccgagcgaactcacggtctattgatggtatg	683
Db	703	TGTTCCGTCAACAGCTTCGAAGGGTTAGAGACACAAAGTTGTGAAGTTGGAGAAATCAGAGG	762
QY	684	agttgatcg-----ggatcgggcccgtttgattccctcgcgctctcttggacggcgagatc	737
Db	763	CGATGAGGGCGAGCTTGATTCGGAGCACTATGATACCATCTGCTATATCTCGAGCGCCGAAATCA	822
QY	738	cctccgaacacgctcttaccggcgcgctcttttcgaaaaatcggaggagaaataaactcgcgtgg	797
Db	823	AAGACGATAAAGGCTACGCGCTCGAGCCGTGAT--GAAGCCGCTCTCGGAGGAGTGTATCG	879
QY	798	agtgggttgacacgaagccgaataattcgggtggtgatatgtctgtcttggggagcgctttga	857
Db	880	AGTGGTTAGACACTAAGCTTGAGCAAGTCGGTGTGTTTGTCTCGTTGTTGCTTCTTGGGA	939
QY	858	ggttccaaaggccaaaatggaaagatctggaaaggcgtattagcctcgggaagccgct	917
Db	940	TCCTCTTTGAGAAGCAACTCGCTGAGTGGCCAAAGGCGTTACAAATCCAACTTTAACT	999
QY	918	ttttatggatgatacgaacacgaagaatgacacgcgcgaagaaagaagaagagtga	977
Db	1000	TCCTGGGTGATCAAGAGAGCTCATATACGCAAGTTACCAAG-----	1044
QY	978	gltgcattggggaaattgaaaaaaatgggaaaaatagtttcgttggtcgcagttgaggg	1037
Db	1045	GGTTTGTGGAAGCTACCAAGACAGAGCGTGTCTGTTCTCTGTGTTAACCGCCTTGAGG	1104
QY	1038	tctctggcgacccctgcgtttggaggttttcgtgacgcattgtgggtgaaactcgcgctg	1097
Db	1105	TTTTAGCTCATGTATCATAGTGTGCTTTTGACTCACTCGCGGTGGAATTCGACGTTGG	1164
QY	1098	agaccttgattgcggggttcccggtgggtgcgcagtcgcttgatcacagacacga	1157
Db	1165	AAGGATTGAGTTTGGGAGTTCCTCATGTTGGTGTGCTCCCTCACTGGAGTGTATCATGAATG	1224
QY	1158	atcgcaagctgattgaggtatgctgggggacagggttgagagtggaaatgaatgaagggg	1217
Db	1225	ATGCTAAGTTGTGGAGGAGGTTTGAGAGTGTGGTATAGAGCGAAGAGGAAGCTGGGG	1284
QY	1218	gtggggttgatgctctgagatagagaggtctgtggagatggtgatggatgggggtgaga	1277
Db	1295	CAGGAGTTGTGAAGACGATGAGGTGGTGCAGGTGTTTGAGAGGAGTGTATGGAAGGAGGA	1344
QY	1278	agagcaaacctagttagagaaaaatgccataaataagaactcttggccagagaagccatgg	1337
Db	1345	GTAGCTGGAGATTAGAGAGAGTTCCTAAGAACTCGAAAGATTTGGCTGTGAAGGGGATCA	1404
QY	1338	gagggatggatcttcaactcaagaatctcaacgcctttct	1377
Db	1405	GTCAAGCAGGAAGCTCTGATCGGAGCATTTAATCAGTTTGT	1444

RESULT	7	
TT10106		
ID	TT10106	standard; cDNA to mRNA; 1731 BP.
AC	TT10106;	
DT	27-OCT-1996	(first entry)
DE	Maize UDP-glucose indol-3-yl:acetyl:glucosyl transferase	
KW	UDP-glucose-indol-3-yl:acetyl:glucosyltransferase; glucanase	
RC	transgenic plant; EC:2.4.1.121; crop improvement; cDNA	
KS	grain; IAGLU-transferase; ds.	
OS	Zea mays.	
FT	Key	Location/Qualifiers
FT	cds	57..1472
FT		/*tag= a
FT		858
FT	misc_binding	/*tag= b
FT		/note= "start of UDP binding region"
PN	W06900291-AL.	
PD	04-JAN-1996.	
PF	20-JUN-1995; U007820.	

24-JUN-1994; US-265427.
(A) RESEARCH CORP TECHNOLOGIES INC.
Handurski RS, Szczygłowski K, Szerzan JB;
WPI: 96-068875/07.
P-PSDB: R39999.
UDP-glucose Indol-3-yl: acetyl:glucosyl transferase DNA and protein
PT - used in sense or anti-sense orientation for the genetic control of
PT plant IAA levels and plant growth
PT Claim 2; page 21-23; 41pp; English
This nucleotide sequence encodes a UDP-glucose-indol-3-
CC ylacetylglucosyltransferase (IAGL-transferase) from maize (Zea mays)
CC Over-expression of this sequence in transgenic plants will result
CC in at least 1 of the following effects: inhibition of apical
CC dominance; inhibition of stem elongation; inhibition of cell
CC enlargement and increased numbers of stems per plant as compared
CC with a wild-type plant. The ability to control the expression of
CC this enzyme allows the control of free indoleacetic acid levels in
CC plants, thereby effecting plant growth rates. Antisense constructs
CC may be used to inhibit synthesis of this enzyme.
CC Sequence 1731 BP; 298 A; 528 C; 601 T;
SQ

Query Match	8.7%	Score 131.8;	DB 1;	Length 1731;
Best Local Similarity	53.7%;	Pred. No. 1.4e-27;		
Matches 324;	Conservative	0;	Mismatches 267;	Indels 12;
				Gaps 2.

Qy	781	ggagataaactcgtggagtgttggacacgaacccgaatacttcgggtggtatgtctc	840
Db	848	GGAGATGGCTGCACCAAGTGGCTAGACACCAAGCCGACCGCTCGTGGCTACGTC	907
Qy	841	gtttggagcgttttgaggtttccaaaggcacaaatggaagagattgggaaaggctatt	900
Db	908	CTTCGCGAGCCTCGCGTCCCTGGCGCAACGCCCCAGAAAGAGAGAGCTCGCGCGCGCCCTCT	967
Qy	901	agctcgcgaaggccgtttttatggatgatacagacacagaagaatgacagcagcggaaga	960
Db	968	CCCCCGCGGCAAGCGCTTCCTGTGGTGTG-----GAGGCCACGACGACGACCA	1018
Qy	961	agaagaagaagattgagttgcattgggaattgaaaaaatgggaaaaatagtttcgtg	1020
Db	1019	GGTCCCGGCTATCTCTCGCCGAGCGGACGCGGCGCGCGATGCTCGTCCGCTG	1078
Qy	1021	gtgtctcagattgagagttctggcgacactcgtgtggatgtttcgtgacgcattgttg	1080
Db	1079	GTGCCCCAGCTGACGTGTGTGGCGACCGGCGCTGGGCTGCTTCGTCAACCCACTGCGG	1138
Qy	1081	gtgaactcggctgtgagagctttagttgcggagttccggtgagcgtgcgcagtg	1140
Db	1139	TGGAATCTCACGCTGAGCGGCTCAGCTTCGGCGTGCCCTATGGTGGCATGGCGTGTG	1198
Qy	1141	gttgatcacagcacgaacatgcaactgattgaggaatcgtggggacaggggtgagagt	1200
Db	1199	GACGGACACCGACCAACGCTCGGAACGTGAGCTGCCCTGGGGGGCGGGGCTGGCGC	1258
Qy	1201	gagaaatgaatgaagg---gggtgggttgatggatctgagatagagaggtgtgtggagat	1257
Db	1259	GCGCCGCGATGTGCGCGGGCGTGTTCCTTCGGGGGAAGTGCAGCGGTGCGTCCGCCG	1318
Qy	1258	ggtgatggtatggggtgagaagagcaaatagtagagagaaatgccataaatggaagac	1317
Db	1319	CGTCAATGACGGGGCGAGGCGGCTGCTGTGCACGAAGGCGGGGGAATGGAGGGA	1378
Qy	1318	tttggccagagaagccatgagagagatggatattcactcaagaatactcaacgccttct	1377
Db	1379	CAGGGCTCGCCCGGGTGGACCGGTGGAGCTCTGACCGCAACCTGGACGAGTTCGT	1438
Qy	1378	tca 1380	
Db	1439	GCA 1441	

RESULT
T66173
8


```

FH Key      Location/Qualifiers
FT CDS      2..1410
FT /tag= a
FT /product= "glucosyl transferase"
FT /note= "encodes Grase from amino acid position 5
FT          shown in W47172"
FT
FT W09745546-A1.
PD 04-DEC-1997.
PD 30-MAY-1997. G01473.
PR 31-MAY-1996; GB-011420.
PA (UTRO-) UNIV YORK.
PI Bowles DJ, Calvert CM, ODonnell PJ, Roberts MR;
DR P-PSDB; W47172.
PT tomato wound inducible (TWIL) gene encoding glucosyl transferase -
PT useful to develop products that alter signalling pathways in plants
PT by altering of salicylic acid, jasmonic acid or ethylene
PS Claim 1; Fig 1; 52pp; English.
CC This wound inducible gene (TWIL) isolated from wounded tomatoes encodes
CC a glucosyl transferase (Grase) protein. This TWIL gene can be used to
CC identify homologue Grase encoding genes isolated from tobacco and rice.
CC A microbial host can be transfected or transformed with a vector
CC containing the Grase encoding nucleic acids. The products can be used to
CC interfere with Grase and therefore alter signalling pathways in plants,
CC specifically tobacco, rice or tomato plants by altering levels of
CC salicylic acid, jasmonic acid or ethylene. This can induce the
CC production of plant defence proteins such as pathogenesis-related (PR)
CC and proteinase inhibitor (PIN) proteins which regulate plant development
CC (plant growth, reproduction and senescence) and improve plant response to
CC pathogens.
SQ Sequence 1624 BP; 545 A; 252 C; 372 G; 455 T;

Query Match      5.0%; Score 76; DB 1; Length 1624;
Best Local Similarity 50.5%; Pred. No. 8.3e-12;
Matches 221; Conservative 0; Mismatches 205; Indels 12; Gaps 1;

QY 779 gagagaataactcgtggagtggttgacacgaagccgaatacttcggtggtatgtg 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 GATGAACACGGCTGCTGAATGCTTGATTCGAGGAATCAAGTTCCATTGTATGTT 838

QY 839 tcgtttggagcgttttgaggtttccaaagccacaataatgggaagagctg 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 TGTTTTGGAGTACAGCAGATTTCTACACGACAGATGCAAGAACTTGTATGGGGCTA 898

QY 899 ttgacctcggagagccgcttttatggatgatacgagacagaagaatgacgcgcaa 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 899 GAAGCCTCTGGACAAAGATTTCAATTTGGGTTATCAGAACAGGGAATGAAGATTGGCTCCA 958

QY 959 gaagagaagaagagttgagttcattgggggaattgaaaaaatgggaataatgttcg 1018
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 959 GAAGGATTCGAGGAAGAAGC-----AAAAGAAAAGGTTTATCATAGAGGA 1006

QY 1019 tgggtctgcagtggtggaggttcggcgcaaccctgcgttgggatttctgtgacgcattgt 1078
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1007 TGGCGACCCCAAGTGTGATTCCTGATCAGGAAGCTATTGGAGCTTTTGTACTCATGTG 1066

QY 1079 gggtagaactcggctgtgagagcttgagttgcgggggttcccggtggtgagcgccag 1138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1067 GGATGGAACTCGACACTGGAAGGAATATCAGCAGGGGGTACCAATGTGACATGCGCCAGTA 1126

QY 1139 tgggtttgacagacgaatcggaagctgattgagagtcgctgggggacagggggtgaga 1198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1127 TTTTGGGAACAGTTTTTCATAGAGAGTGTGCTGACTGAGTGAATGAGAAGTGGAGCTGTT 1186

QY 1199 gtgagaatgaatgaagg 1216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1187 GTTGTCTTAAACCAATGG 1204

RESULT 11
X24873
ID X24873 .standard; cDNA; 1626 BP.
```

```

AC X24873;
DT 21-JUN-1999 (first entry)
DE Medium-chain UDP-glucose:fatty acid glucosyltransferase cDNA.
KW UDP-glucose:fatty acid glucosyltransferase; tomato;
KW 1-O-acyl-a-glucose; acyl donor; esterification; ss.
OS Lycopersicon pennellii.
OS Lycopersicon esculentum.
FH Key      Location/Qualifiers
FT CDS      1..1416
FT /tag= a
FT
FT W09909144-A1.
PD 25-FEB-1999.
PD 10-AUG-1998; IB1369.
PR 30-JUN-1998; US-106464.
PR 13-AUG-1997; US-055554.
PA (CORR ) CORNELL RES FOUND INC.
PI Eanneatta NT, Ghangas GS, Kuai J, Steffens JC;
DR WPI; 99-190155/16.
DR P-PSDB; W98009.
PT New chain length specific uridine diphosphate-glucose:fatty acid
PT glucosyltransferases - useful for preparing straight or branched,
PT (unsaturated 1-O-acyl-a-glucoses and 1-O-acyl-a-D-glucoses
PT Example 3; Page 33-36; 42pp; English.
PS The present sequence encodes a UDP-glucose:fatty acid
CC glucosyltransferase that transfers glucose to medium chain length
CC fatty acids to form 1-O-acyl-a-glucoses. The 1-O-acyl-a-glucoses
CC act as acyl donors in the esterification of glucose and in the
CC further esterification of partially acylated glucose in reactions
CC catalysed by glucose acyltransferases. They also act as acyl donors
CC in the esterification of sucrose and in the further esterification
CC of partially acylated sucrose in reactions catalysed by sucrose
CC acyltransferases. The source of the present cDNA is leaf trichome
CC mRNA of an F1 population of a cross between wild tomato
CC Lycopersicon pennellii (LA 716) and cultivated tomato
CC Lycopersicon esculentum cv. New Yorker. The clone was isolated
CC following immunoscreening of a trichome cDNA library. The
CC invention provides chain length-specific UDP-glucose:fatty acid
CC glucosyltransferases and a method for their purification. The
CC enzymes have specificity for transferring glucose to short, medium
CC or long chain length fatty acids. Methods for preparing
CC 1-O-acyl-a-D-glucoses, where the acyl group comprises 3-5, 6-13
CC or 14-22 C atoms, are claimed.
SQ Sequence 1626 BP; 540 A; 256 C; 372 G; 458 T;

Query Match      5.0%; Score 76; DB 1; Length 1626;
Best Local Similarity 50.5%; Pred. No. 8.3e-12;
Matches 221; Conservative 0; Mismatches 205; Indels 12; Gaps 1;

QY 779 gagagaataactcgtggagtggttgacacgaagccgaatacttcggtggtatgtg 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 793 GATGAACACGGCTGCTGAATGCTTGATTCGAAGAAATCAAGTTCCATTGTATGTT 852

QY 839 tcgtttggagcgttttgaggtttccaaagccacaataatgggaagagattgggaaggcta 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 853 TGTTTTGGAGTACAGCAGATTTCTACACGACAGATGCAAGAACTTGTATGGGCTA 912

QY 899 ttgacctcggagagccgcttttatggatgatacgagacagaagaatgacgcgcaa 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 913 GAAGCCTCTGGACAAAGATTTCAATTTGGGTTATCAGAACAGGGAATGAAGATTGGCTCCA 972

QY 959 gaagagaagaagagtggtgagttgcattggggaattgaaaaaatgggaataatgttcg 1018
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 973 GAAGGATTCGAGGAAGAAGC-----AAAAGAAAAGGTTTATCATAGAGGA 1020

QY 1019 tgggtctgcagtggtggaggttcctggcgcaaccctgcgttgggatttctgtgacgcattgt 1078
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 TGGCGACCCCAAGTGTGATTCCTGATCAGGAAGCTATTGGAGCTTTTGTACTCATGTG 1080

QY 1079 gggtagaactcggctgtgagagcttgagttgcgggggttcccggtggtgagcgccag 1138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1081 GGATGGAACCTCGACACTGGAAGGAATATCAGCAGGGGTACCAATGTTGACATGGCCAGTA 1140
```

```
QY 1139 tgggttgatcagacagcaaatggaagctgattgagagatcgctgggggacaggggtgaga 1198
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1141 TTTCGGGAACACTTTTTCATGAGAAGTTGGTGACTAGCTAATGAGAAAGTGGAGCTGGT 1200
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1199 gtgagaaatgaatgaagg 1216
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1201 GTTGGTTCTAACAATGG 1218
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
T68693
ID T68693 standard; cDNA; 432 BP.
AC T68693;
AT 05-JAN-1998 (first entry)
DE Strawberry UDP-glucuronosyl transferase cDNA.
KW Strawberry; UDP-glucuronosyl transferase; ERRLb; fruit; ripening;
KW storage; gene regulation; ss.
OS Fragaria ananassa Duch. cv. Brighton.
PN W09721816-A1.
PD 19-JUN-1997.
PF 12-DEC-1996; G03076.
PR 13-DEC-1995; GB-025459.
PA (ZENE ) ZENECA LTD.
PI Manning K.
DI WPI: 97-332787/30.
PT Vectors encoding new enzymes for regulating ripening of fruit - used
PT particularly to improve storage properties, processing
PT characteristics etc, especially in strawberries
PS Claim 1; Page 23-24; 43pp; English.
CC This cDNA sequence comprises a UDP-glucuronosyl transferase (ERT
CC 1b) cDNA clone isolated from a ripe strawberry fruit cDNA library
CC by differential screening. 9 Ripening-enhanced clones (see T68687-
CC 95) clones were isolated from the library, and are deposited at
CC NCIMB 40693. Any of these 9 sequences can be used as a regulatory
CC sequence in claimed vectors for genetic transformation of plant
CC cells to regulate fruit ripening. The vectors also comprise a
CC promoter and transcription termination sequence. Also claimed are
CC plants (especially strawberry) and propagating materials containing
CC these vectors. Stable integration of the regulatory sequences, or
CC their complements or genomic equivalents, into a plant is used to
CC modulate fruit ripening by overexpression or downregulation of an
CC endogenous plant gene. Slowing ripening of fruit improves
CC resistance to damage during harvesting, packaging and transport.
CC It also extends shelf life, improves storage properties, processing
CC characteristics, flavour and aroma, and modifies colour and
CC increases resistance to post-harvest fungi.
SQ Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;

Query Match 4.8%; Score 72.6; DB 1; Length 432;
Best Local Similarity 59.4%; Pred. No. 3.9e-11;
Matches 123; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 997 aaaaatgggaaaaatgttcgtgctgcagttgaggttctggtgcgcacccctgcgtt 1056
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 AGACAGAGCAAAAGTTGTGCAATGGAGTCCACAGAGAAGATTTTGAGCATCCTTCGAC 164
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1057 gggatgttcgtgagcattgtgggtggaactcggctgtgagagcttgagttcggggt 1116
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 GGCTTCGTTGTGACTCATTCGGGGTGGAACTCAACCACTGAGTCACTACCCFCAGGAAT 224
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1117 tccgggtgggtggcgcagtggtttgatcagacgacgaatcgaaagctgattgagga 1176
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 GCCCGTGGTGGCATTCACCAATAGGGGTGACCAAGTACCAGACGCCCAAGTATTGGTCCA 284
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1177 tgcgtgggggacaggggtgagagtgag 1203
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 CGAGTTTAAGGTGGGAGTAAGAATGTG 311
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
V23108
ID V23108 standard; cDNA to mRNA; 1597 BP.
```

```
AC V23108;
DT 06-AUG-1998 (first entry)
DE UDP-glucose:flavonoid 3,5-O-glucosyl transferase gene.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
KW blue flower; ds.
OS Unidentified.
FH Key Location/Qualifiers
FT CDS 39..1460
      /*tag= a
FT J10113184-A.
PN 06-MAY-1998.
PD 06-MAY-1998.
PF 15-OCT-1996; 272253.
PR 15-OCT-1996; JP-272253.
PA (TOLG ) TOKYO GAS CO LTD.
DR WPI: 98-315473/28.
DR P-PSDB; W56451.
PT Gentian flavonoid glucosyl transferase gene - used to produce
PT recombinant rose plants with blue petal(s)
PS Claim 3; Pages 7-8; 10pp; Japanese.
CC The present sequence encodes a UDP-glucose:flavonoid 3,5-O-glucosyl
CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (sic) petals. The gene is
CC used to transform rose plants to produce blue flowers.
SQ Sequence 1597 BP; 483 A; 272 C; 368 G; 474 T;

Query Match 4.6%; Score 69.4; DB 1; Length 1597;
Best Local Similarity 57.7%; Pred. No. 6e-10;
Matches 124; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 998 aaaaatgggaaaaatgttcgtggtgcagttgaggttctggtgcgcacccctgcgttg 1057
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035 AACACCGGAATGGTATGGATGGGCCCCACAGACGGCCGTGTGTCTCACTCAGCTGTG 1094
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1058 ggaatgttcgtgacgcattgtgggtggaaactcggctgtgagagcttgagttcggggtt 1117
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1095 GGAGGTTTCGTCTCGCATTTGGATGGGAATTCACACTGGAAACATTTGGTTGGTGTC 1154
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1118 ccggtgtggtggtggtgcgcagttggttgcagcagcgaatcgaaagctgaggtgaggt 1177
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1155 CCAGTGGCAACCTGGCCCTCTGTTCTGTCAGCAGCAGATGATGATTTGAATTGGTGAG 1214
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1178 gcgtgggggacaggggtgagagtgagagtgagtgagtgagtgagtgagtgagtgagtgag 1212
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1215 GAATGGGCTTGGCTGTGGAGGTTAAGATGGATTA 1249
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
T62123
ID T62123 standard; cDNA to mRNA; 1578 BP.
AC T62123;
DT 10-JUN-1997 (first entry)
DE Solanum melongena flavanoid-3-glucosyl transferase cDNA.
KW Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;
KW pigment; production; seed; plant; ds.
OS Solanum melongena.
FH Key Location/Qualifiers
FT CDS 13..1314
      /*tag= a
      /product= flavanoid-3-glucosyl_transferase
FT J09056385-A.
PN 04-MAR-1997.
PD 04-MAR-1997.
PF 25-AUG-1995; 238943.
PR 25-AUG-1995; JP-238943.
PA (KIRI ) KIRIN BREWERY KK.
DR WPI: 97-208631/19.
DR P-PSDB; W13406.
PT Solanum melongena flavonoid-3-glucosyl transferase and DNA encoding
PT it - catalyses transfer of glucose to 3-position of flavonoid
PT pigment in plants
PS Claim 2; Page 10; 14pp; Japanese.
CC The present sequence encodes Solanum melongena (eggplant)
CC flavanoid-3-glucosyl transferase (F3GT), which catalyses the
```

```
CC transfer of glucose to the 3-position of flavonoid pigments. The
CC F3GT DNA is useful for the production of the seed of a plant. A
CC S. melongena seed was cultured under a red light and a flavanoid
CC derived. A cDNA library in which the pigment was concentrated was
CC prepared. A flavanoid glucosidising enzyme gene was cloned and its
CC structure and mRNA expression pattern analysed. The clone was
CC expressed in E. coli.
SQ Sequence 1578 BP; 498 A; 264 C; 310 G; 506 T;

Query Match 4.6%; Score 69.2; DB 1; Length 1578;
Best Local Similarity 49.2%; Pred. No. 6.8e-10;
Matches 317; Conservative 0; Mismatches 288; Indels 29; Gaps 4;

QY 773 aaatcgagagagaataactcgtgaggtgttgagcagcagcagcgaatacttcggtggtg 832
DB 736 AAATAGATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 795
QY 833 tatgtcgttggagcgttttgaggtttccaaaggccaaatggaagagatggaaa 892
DB 736 TATCTTATGTTTGGCACTGTAAACAACATTACCCCTTAACGAATTTGGTGTGATAGCAGAA 855
QY 893 gggctattagcctcgagagcgttttatgtatgatacagcagcagcagaatgacgac 952
DB 856 GCCTTAGAAGTAAAGAAACCCCTTTATTTGGTCTAATTAAGAAATTAAGTGTCAAGAT 915
QY 953 ggcgaagaagaagaagaggtgagttgattcattgggggaattgaaaaaattgggaaata 1012
DB 916 TTGCCTAAAG-----GATTTCTTGAAGAACAAAGAGAGTTTGGGAAATA 960
QY 1013 gtttcgtggtgctcgaggttgaggttctggcgcacccctgctggtggatgttttcgtgac 1072
DB 961 GTTCTTGGGCACCCCAATTTGGAATCTTGGCACATAGTCTGTGTGGTGTTCGTAAACG 1020
QY 1073 cattgtgggtggaactgctgtggagagcttgagttgcggggttcccggtggtggtg 1132
DB 1021 CATTTCCGGATGGAATTCGATTTTGGGAAGGCATTTCAATTTGGTGTGCTTATGATATGAG 1080
QY 1133 ccgcagtggtttgatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1192
DB 1081 CCTTTTGTGTGACCAAAATTTGAATAGTAGAATGGTGAAGGTGTTTGGGAATTTGCT 1140
QY 1193 gtgagagtgagaaatgaatgaaggggtggtggttgaatgagatcagagagagtggtg 1252
DB 1141 TTGCAA-----ATTGAAGGTGGGATTTTACTAAAGTGGGAATTAATTAGTCTTTG 1191
QY 1253 gagatggtgagtggtggtgagagagcagcagcagcagcagcagcagcagcagcagcagc 1312
DB 1192 GATACTTTTCAATAG---GAAAAAGGGAAGATATTAAAGGGAATTTGAAGGCTA 1248
QY 1313 aagactttggcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1370
DB 1249 AAAGAAAAGCATTTAGAAAGTGTGAACACAGATTTGAAGAGTTTGAACAGAAAATTTCAAG 1308
QY 1371 cctttcttcaatgaagttgcagcgtgcttaatacacaaaaatggctt 1414
DB 1309 TTCTAGTTGAGCTAGTTAAATGTGCACAAAGTCAACTTAATATGTT 1352

RESULT 15
Q74684
ID Q74684 standard; cDNA; 1650 BP.
AC Q74684;
DT 22-JUN-1995 (first entry)
DE Early Ripening Tomato protein ER1b gene.
KW Early Ripening Tomato; transgenic plant; crop improvement;
KW fruit ripening; transformation; ss.
OS Lycopersicon esculentum.
PN W09421794-A.
PD 29-SEP-1994.
PF 22-MAR-1994; G00581.
PR 22-MAR-1993; GB-005860.
PR 22-MAR-1993; GB-005862.
```

```
PR 22-MAR-1993; GB-005865.
PR 22-MAR-1993; GB-005869.
PR 22-MAR-1993; GB-005866.
PR 22-MAR-1993; GB-005869.
PR 22-MAR-1993; GB-005867.
PR 22-MAR-1993; GB-005868.
PR 12-JUL-1993; GB-014351.
PR 12-OCT-1993; GB-020988.
PA (ZENEC) ZENEC LTD.
PI Abu-Bakar UK, Barton SL, Gallego-Velgas PP, Gray JE;
PI Grierson D, Lowe AL, Picton S, Whotton LC, Abubakar UK;
DR WPI; 94-317014/39.
PT DNA constructs encoding fruit-ripening related proteins - useful
PT for transformation of plants to modify fruit quality
PS Claim 1: Page 47: 74pp: English.
CC ER1b homologous transcripts are 1.8 kb in size and it is only
CC expressed during ripening of the wild-type fruit. Its expression is
CC highest during early stages of fruit ripening (e.g. breaker plus 3).
CC Levels of expression of the mRNA are low in rin (ripening inhibitor)
CC mutant tomatoes and are restricted to ripening rin tomatoes. The
CC gene is not activating upon wounding. The sequence has been
CC deposited as NCIMB 40544. The DNA may be used to transform plants
CC for production of plants with an altered level of expression.
CC Improvements in transgenic plants include resistance to damage and
CC pests, longer shelf life, improved flavor/aroma, etc..
SQ Sequence 1650 BP; 492 A; 338 C; 348 G; 472 T;

Query Match 4.5%; Score 68.4; DB 1; Length 1650;
Best Local Similarity 50.3%; Pred. No. 1.2e-09;
Matches 196; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

QY 779 gagagaataactcgtgaggtgttgacacagcagcagcagcagcagcagcagcagcagcagc 838
DB 732 GAACAAATAGCTGATCGAGTGGCTGATACACAAAGCCCTTAACCTCTGTCTCTATGTC 791
QY 839 tcgtttgggagcgttttgaggtttccaaaggcacaatggaagagattgggaaggcgcta 898
DB 792 AGCTTGGTAGCTAGTGTAGGATTTGATCACAAGAGCTTGTGACACTCTTGGGATTA 851
QY 899 ttgacctgcggaaggccggttttatgtatgatacagcagcagcagcagcagcagcagcagc 958
DB 852 GCTAATAGCGATCAACCGTCTTGTGGGTATTTCGACCTTGGCTCTGTCTCTGCTTTCAA 911
QY 959 gaagaagaagaagagttgagttgattggggaattgaaaa---aaatgggggaaatagtt 1015
DB 912 TGTGCTGAGGCACCTGCTGATGTTTGTGAGAAAATGGTAGGAGAAAGAGACGAATAGTG 971
QY 1016 tcgtggtgctcaggttgaggttctgcgcacccctgcttgggtggtggttcttcgacgcat 1075
DB 972 AAATGGGCAACCAAAAACAGGTGTTGCACATCCCGGTAGCAGAGTTTTCACATCAT 1031
QY 1076 tgtgggtggaactcggctgtgagagagcttgagttgcggggttcccggtggtggtggtg 1135
DB 1032 TGTGTTGGAATTTACGCTTGAAGTATATGTGAAGAAGTCCCTATGTGTGTGTCAGGCCA 1091
QY 1136 cagtggtttgatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1165
DB 1092 TTTCTAGCAGACCAACTGGTGAACGCAAGG 1121

Search completed: August 1, 2000, 18:16:53
Job time: 19934 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2000, 21:39:11 ; Search time 152.01 seconds
(without alignments)
1363.704 Million cell updates/sec

Title: US-09-147-955-1

Perfect score: 1507
Sequence: 1 gaaatttcacaaaaattg.....ttataatttagtaacaaaa 1507

Scoring Table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
- 7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.8	8.7	1731	4	US-08-466-583-1
2	131.8	8.7	1731	6	PCT-US95-07820-1
3	84.2	5.6	7218	1	US-08-232-463-14
4	76	5.0	1627	5	US-08-106-464-1
5	66.8	4.4	1669	3	US-08-522-421-1
6	56.2	3.7	1607	4	US-08-797-226-1
7	55.2	3.7	405	4	US-08-975-316-32
8	54.4	3.6	1738	3	US-08-379-482A-2
9	46.6	3.1	28958	1	US-08-258-261B-6
10	46.6	3.1	28958	1	US-08-456-837-6
11	46.6	3.1	28958	1	US-08-457-342-6
12	46.6	3.1	28958	1	US-08-457-046A-6
13	46.6	3.1	28958	2	US-08-458-076A-6
14	46.6	3.1	28958	2	US-08-764-233A-4
15	46.6	3.1	28958	2	US-08-457-335A-6
16	46.6	3.1	28958	2	US-08-729-214-6
17	46.6	3.1	49377	2	US-08-764-233A-1
18	45.2	3.0	13987	3	US-08-804-227C-13
19	45.2	3.0	44377	3	US-08-804-227C-7
20	45.2	3.0	44377	4	US-08-804-198-1
21	44.6	3.0	1505	1	US-07-915-246-1
22	44.6	3.0	2336	6	PCT-US92-00282-1
23	44.6	3.0	2339	6	PCT-US92-00282-2
24	43.6	2.9	1358	2	US-08-471-033-45
25	43.6	2.9	1358	3	US-08-471-044-45
26	43.6	2.9	1358	3	US-08-463-483A-45

27	43.6	2.9	1358	3	US-08-471-046A-45	Sequence 45, Appl
28	43.6	2.9	1358	3	US-08-470-566B-45	Sequence 45, Appl
29	43.6	2.9	1358	4	US-08-469-334-45	Sequence 45, Appl
30	43.6	2.9	1358	5	US-09-300-529-45	Sequence 45, Appl
31	43.4	2.9	1338	3	US-08-484-993B-7	Sequence 7, Appl
32	43.4	2.9	1338	4	US-08-484-158B-7	Sequence 7, Appl
33	43.4	2.9	1338	4	US-08-484-596A-7	Sequence 7, Appl
34	43.4	2.9	1338	4	US-08-480-150A-7	Sequence 7, Appl
35	43.4	2.9	1338	5	US-08-458-731-7	Sequence 7, Appl
36	43.4	2.9	1338	5	US-08-149-223A-7	Sequence 7, Appl
37	42.8	2.8	2064	1	US-08-343-428-1	Sequence 1, Appl
38	42.2	2.8	2589	4	US-08-482-728A-3	Sequence 3, Appl
39	41.6	2.8	4897	7	5196516-7	Patent No. 5196516
40	41.6	2.8	53526	5	US-08-658-136-2	Sequence 2, Appl
41	41.6	2.8	53577	5	US-08-658-136-1	Sequence 1, Appl
42	41.4	2.7	434	3	US-08-332-766A-10	Sequence 10, Appl
43	41	2.7	450	4	US-08-387-942C-40	Sequence 40, Appl
44	41	2.7	12588	4	US-08-387-942C-1	Sequence 1, Appl
45	40	2.7	9960	5	US-08-822-586-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-08-466-583-1
; Sequence 1, Application US/08466583
; Patent No. 591998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..1472
US-08-466-583-1

Query Match 8.7%; Score 131.8; DB 4; Length 1731;

Best Local Similarity 53.7%; Pred. No. 8.5e-27;

Matches 324; Conservative 0; Mismatches 267; Indels 12; Gaps 2;

```
QY 781 ggagataaactcgtgagtggttgacacgaagcgaatcttcgtggtggtatgtgc 840
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 848 GGAGATGCGTGACCAAGTGGCTAGACACCAAGCCCGACCCCTCGGTGGCTACGTC 907
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 gtttgagagcgttttgaggtttccaaaggccacaaatgggaagattgggaaggctatt 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 908 CTTCCGACGCTCGCTCCCTGGSCAACGCCAGAGAGGAGCTCGCGCGCGCTCCT 967
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 agctcgcgaagccgtttttatggatgacagacagaagaatgacagcgcgaaga 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 968 CGCGCCGGCAAGCCGTCTCTGGTGGT-----GAGGCCACGACGACACCA 1018
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 agaagaagaagattgagttgcatctggaatggaataatgggaataatgatttcgtg 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1019 GGTCCCGCGCTATCTCTGGCCGAGCGACGCGCGCGCGCGCTGCTGCTGCGCTG 1078
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 gtgctcgcagttgaggttctgacacacctgctggtggaatggttcgtaacacattg 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1079 GTGCCCCGACGTGGACGTGTGGCGCACCGCGCTGGCTGCTGCTACCCACTGGCG 1138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 gtgaactcgcgtgagagcgttgagttccaaaggccacaaatgggaataatgatttcgtg 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1139 TTGGAATCCACGCTGGAGCGCTCAGCTTCGCGCTGCCCTATGCTGGCTGCTG 1198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1141 gtttgatcagacacgaatcgaagctgattgaggtgaggtgaggtgaggtgaggt 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1199 GACGGACACGCGCGCTGCTGGCGCACCGCGCTGGCGCGCGCGCTGCTGCTGCTG 1258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1201 gagaaatgaagg---gggtgggttgatgagatctgagatgagaggtggtgagat 1257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1259 GCGCGCGATGCTGGCGCGCGCTGTCTTCGCGGGAAGTGGAGCGGTGCTGCGCG 1318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1258 gttgatgaggtgagagacaaactagtgagagaaatgccataaataatgaaagac 1317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1319 CGTCATGACGGGGCGCGCGCTGCTGCACCAAGCGCGCGGGGGAATGAGGGA 1378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1318 ttggccagagaagcattggagaggtgatttctcactcaagaatctcaagcctttct 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1379 CAGGCTCGCGCGCGGTGGCACCGCGTGGACCTGTGACCGAACCTGGACAGTTCGT 1438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1378 tca 1380
    ||
DB 1439 GCA 1441
    ||
```

RESULT 2

PCT-US95-07820-1

Sequence 1, Application PC/TUS9507820

GENERAL INFORMATION:

APPLICANT: Bandurski, Robert S.

APPLICANT: Szafer, Jozef B.

APPLICANT: Szafer, Jozef B.

TITLE OF INVENTION: Genetic Control of Plant Hormone Levels

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Greenlee & Winner, P.C.

CITY: Boulder

COUNTRY: United States of America

STATE: Colorado

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07820

FILING DATE: 19-JUN-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/265,427

FILING DATE: 24-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 11-94B PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/499-8080

TELEFAX: 303/499-8089

TELEX: 49617824

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1731 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA to mRNA

FEATURE:

NAME/KEY: CDS

LOCATION: 57..1472

PCT-US95-07820-1

Query Match

Best Local Similarity 53.7%; Pred. No. 8.5e-27;

Matches 324; Conservative 0; Mismatches 267; Indels 12; Gaps 2;

```
QY 781 ggagataaactcgtgagtggttgacacgaagcgaatcttcgtggtggtatgtgc 840
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 848 GGAGATGCGTGACCAAGTGGCTAGACACCAAGCCCGACCCCTCGGTGGCTACGTC 907
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 gtttgagagcgttttgaggtttccaaaggccacaaatgggaagattgggaaggctatt 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 908 CTTCCGACGCTCGCTCCCTGGSCAACGCCAGAGAGGAGCTCGCGCGCGCTCCT 967
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 agctcgcgaagccgtttttatggatgacagacagaagaatgacagcgcgaaga 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 968 CGCGCCGGCAAGCCGTCTCTGGTGGT-----GAGGCCACGACGACACCA 1018
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 agaagaagaagattgagttgcatctggaatggaataatgggaataatgatttcgtg 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1019 GGTCCCGCGCTATCTCTGGCCGAGCGACGCGCGCGCGCTGCTGCTGCGCTG 1078
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 gtgctcgcagttgaggttctgacacacctgctggtggaatggttcgtaacacattg 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1079 GTGCCCCGACGTGGACGTGTGGCGCACCGCGCTGGCTGCTGCTACCCACTGGCG 1138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 gtgaactcgcgtgagagcgttgagttccaaaggccacaaatgggaataatgatttcgtg 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1139 TTGGAATCCACGCTGGAGCGCTCAGCTTCGCGCTGCCCTATGCTGGCTGCTG 1198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1141 gtttgatcagacacgaatcgaagctgattgaggtgaggtgaggtgaggtgaggt 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1199 GACGGACACGCGCGCTGCTGGCGCACCGCGCTGGCGCGCGCGCTGCTGCTGCTG 1258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1201 gagaaatgaagg---gggtgggttgatgagatctgagatgagaggtggtgagat 1257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1259 GCGCGCGATGCTGGCGCGCGCTGTCTTCGCGGGAAGTGGAGCGGTGCTGCGCG 1318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1258 gttgatgaggtgagagacaaactagtgagagaaatgccataaataatgaaagac 1317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1319 CGTCATGACGGGGCGCGCGCTGCTGCACCAAGCGCGCGGGGGAATGAGGGA 1378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1318 ttggccagagaagcattggagaggtgatttctcactcaagaatctcaagcctttct 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1379 CAGGCTCGCGCGCGGTGGCACCGCGTGGACCTGTGACCGAACCTGGACAGTTCGT 1438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1378 tca 1380
    ||
```


Db	793	GATGAACACGCGTCTTGAAATGGCTTGATTCGAAGAAATCAAGTTCATTTCCATTTGTTATGTT	852
QY	839	tcggttgaggagcgttttgaggtttcccaagcgacacaaatgaaagagattgggaagggccta	898
Db	853	TGTTTTGGAAGTACAGCAGATTTCTACTACAGCAGATGTCAGAACTTCTGTATGGGGCTA	912
QY	899	ttagcctcgcgaagccggtttttatgtagatcacgaaacagaagaatagaacgcgcgcaa	958
Db	913	GAAGCCTCTGCAGCAAGATTTTCATTTTGGGTTATCAGAACAGGGAAATCAAGATTGGCTCCCA	972
QY	959	gaagaagaagaagtgtagttgcattcggggaattgaaaaaatggggaatatgttcg	1018
Db	973	GAAGCTTTCGAGGAAGAAGC-----AAAAGAAAAGGTTTAAATCATAGAGGA	1020
QY	1019	tggctctcgagttgaggttctgcgcacccctgcgttgggaatggttctgcgcacattgt	1078
Db	1021	TGGGACCCCAAGTGCTGATTTCTGATCAGAAAGCTATTGGAGCTTTTGTACTCATTTGT	1080
QY	1079	gggtggaactcgcgtgtgagagccttgagttgcggggttcccggtggtgcgctgcgcag	1138
Db	1081	GGATGAACTCGACACTGGAAGGAATATCAGACGGGTACCAATGTTGACATGGCCAGTA	1140
QY	1139	tggtttgatcagaacgcgaatcgaagctgattgagatcgcgtggggggacaggggtgaga	1198
Db	1141	TTTGGCGAAAGAGTTTTTTCATGAGAAAGTTGGTGACTGAGGTAAATGAGAAAGTGGAGCTGGT	1200
QY	1199	gtgagaatgaatgaagg	1216
Db	1201	GTTCGTTCTTAAGCAATGG	1218

RESULT 5

US-08-522-421-1
; Sequence 1, Application US/08522421

```

: Patent No. 5908973
:
: GENERAL INFORMATION:
:
: APPLICANT: Abu-Bakar, Uni Kalsom
: APPLICANT: Barton, Sarah Louise
: APPLICANT: Gallego-Veigas, Pedro Pablo
: APPLICANT: Gray, Julie Elizabeth
: APPLICANT: Grierson, Donald
: APPLICANT: Lowe, Alexandra Louise
: APPLICANT: Picton, Steve Colin
: APPLICANT: Whotton, Lee Colin
:
: TITLE OF INVENTION: DNA, DNA CONSTRUCTS,
: TITLE OF INVENTION: DERIVED THEREFROM
:
: NUMBER OF SEQUENCES: 10
:
: CORRESPONDENCE ADDRESS:
:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305865.9
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305866.7
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305867.5
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305860.0
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305862.6
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9314351.9
; FILING DATE: 12-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9320988.0
; FILING DATE: 12-OCT-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: ERTL1
US-08-522-421-1

```

Query Match 4.4%; Score 66.8; DB 3; Length 1669;
Best Local Similarity 50.0%; Pred. NO. 5.5e-09;
Matches 195; Conservative 0; Mismatches 192; Indels 3

Qy	779	gaggaagaataactgcgtggagtggtgacacgaagccgaatacttcgggtgggtgatcgtg	838
Db	732	GAAGACAATAGCTGCATCGAGTGGGTGCGATAGACAAGCCCTAACTGTGTTCTCTATGTC	791
Qy	839	tcggtttgggagcgttttgaggtttccaaagggcacaaaTggagagattgggaagaggcta	898
Db	792	AGCTTGGGTAGCCTAGTAGGAGTTGATGATCAAAAGAGTTGATTGAGACTGCTTGGCGATTA	851
Qy	899	ttagcctcggaagccggtttttatgagtgtacagagaacagaagaatgacgacgcgcaa	958
Db	852	GCTAATAGCATCAACCGTCTCTGTGGGTATTCTGCACCTGGCTCTGTCTCTGGCTTTCAA	911
Qy	959	gaagagaagaagaagtttgagttgcatttgggggaattgaaaa--aaatcggggaaaaatagtt	1015
Db	912	TGTGCTGAGGCATGCCTGATGGTTTGTGAAAAATGGTAGGAGAAAGAGGACGAATAGTG	971
Qy	1016	tcgtgggtgcgcagtttgaggtttctggcgacacctcgcttgggaatgtttcgtgcagccat	1075
Db	972	AAATGGGCACCAAAAACAGGTGCTGCACATCCGCGGTAGCAGGGTTTTTTCACTCAT	1031
Qy	1076	tgtgggtgaaactcggctgtggagagcttgagttcgggggttcctcgggtggtggcggtccg	1135
Db	1032	TGTGGTTGGAAATTTCTACGCTTGAAGATATATGTGAAGAAGTCCCTATGTTGTGCGAGGCCA	1091
Qy	1136	cagtggtttgatcacgacgacgaatcgaa	1165
Db	1092	TTTCTAGCAGACCAACTGGTGAACGCAAGG	1121

RESULT

RESULTS 6
US-08-797-226-1
; Sequence 1, Application US/08/97226
; Patent No. 5959180
; GENERAL INFORMATION:
; APPLICANT: MOEHS, CHARLES P
; APPLICANT: ALLEN, PAUL V

APPLICANT: ROCKHOLD, DAVID R	QY	795	tggagttggttgacaaagaacccgaaaatottcggtggtgtatgtgctgtttgggagcgcttt	854
APPLICANT: STAPLETON, ANDREW				
APPLICANT: GABBARINO, JOAN E	DB	852	TAGATTGTTTGAATGTCACAGAAACCTTAAATCGTTCCTATCTATCTTTCGGAAGCATGG	911
APPLICANT: FRIEDMAN, MENDEL				
APPLICANT: BELKNAP, WILLIAM R	QY	855	tgaggtttccaaaagggcacaaatggaagagattgggaagaggtctattagcctgcggaagc	914
TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE				
TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE	DB	912	CTAGATTTCCTGAGAGCCAACTCAATGAATAAGCCCAAGCTCTGGATGCTTCAAAATGTC	971
NUMBER OF SEQUENCES: 2				
CORRESPONDENCE ADDRESS:	QY	915	cgtttttatggatgatacagagaacagaagaatagcagacgcggaagaagaagaagaagagt	974
ADDRESSEE: NANCY J. PARSONS				
STREET: 800 BUCHANAN ST.	DB	972	CTTTTCATTTTGTATTGAG-----GCCCTAATGAAGAAACGGCGCTGGTGGTGCAGTTG	1025
CITY: ALBANY				
STATE: CA	QY	975	tgagttgcattgggggaattgaaaaaaatggggaaaaatagtttcgtggtgctgcagctgg	1034
COUNTRY: USA	DB	1026	GTAATTTTAGAGGACACACTAAAAGGGTTTGTACATCAAAAGGGTGGGTCCACAGCTTA	1085
ZIP: 94710				
COMPUTER READABLE FORM:	QY	1035	aggcttcggcgaccctgcgttgggatgtttcgtgacgcatgtgggtgggagacccgctg	1094
MEDIUM TYPE: Floppy disk	DB	1086	CGATCATGGAAACATTCAGCAACAGCGGGTTCATGACTCATTTGTGTACTAATTCGGTTC	1145
COMPUTER: IBM PC compatible				
OPERATING SYSTEM: PC-DOS/MS-DOS	QY	1095	tggagagcttgagttgcgggggttccggtggtggtggtgcgcagtggtttgatcagacga	1154
SOFTWARE: PatentIn Release #1.0, Version #1.25	DB	1146	TGGAAGCCATCACCTTTTGGCGTGCCAAATACATGATGATGGCCACTTTATGCTGATCAATCT	1205
CURRENT APPLICATION DATA:	QY	1155	cgaaatgcgaagctgattggagatgcgtggggg-----acaggggtgagatgagaatga	1208
APPLICATION NUMBER: US/08/797,226	DB	1206	ACACGAGNAGGTAGTCGAGGTAGGGGATTGGGAATCAAAATCGGGATAGATGTATGGA	1265
FILING DATE:				
CLASSIFICATION: 536	QY	1209	atgaaggggtgggggttgatggatcgtgatatgagag	1245
ATTORNEY/AGENT INFORMATION:	DB			
NAME: PARSONS, NANCY J				
REGISTRATION NUMBER: 40,364	QY			
REFERENCE/DOCKET NUMBER: 0011.97	DB	1266	ATGAAGGGATTGAGATCAGCGGCCCTGTATATGAAAG	1302
TELECOMMUNICATION INFORMATION:				
TELEPHONE: (510) 559-5731				
TELEFAX: (510) 559-5777				

Query Match 3.78; Score 56.2; DB 4; Length 1607;
Best Local Similarity 48.69; Pred. No. 4.3e-06;
Matches 222; Conservative 0; Mismatches 223; Indels 12

RESULT 7
US-08-975-316-32
; Sequence 32, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKBERG, Leonard N.; HAVUKKALA, Ilkka
; APPLICANT: AND GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SLEATH, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:


```

; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8889
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
US-08-258-261B-6

Query Match          3.1%; Score 46.6; DB 1; Length 28958;
Best Local Similarity 47.4%; Pred.No.0.0078;
Matches 176; Conservative 0; Mismatches 189; Indels 6; Gaps 1;

QY 144 ggcgcgcatggccaacacagcctccgccgtccggaaaccacccggcgctcgacttcg 203
DB 10208 GCACAGCCAGGGCGGATGCCGCCCTTCGCGCAGCGCTCTCTCCCTCGAGGACG 10267

QY 204 tggcgttctccgacgggttacgacgacgggtgaagccctcgcgcgacgggaagcgctaca 263
DB 10268 CCGCGCGCATGCCGCCCTTGCAGGAAAGCGGTCAACACGTCGGCGCAACGGCGGA 10327

QY 264 tgtccgagtgaagccgcgcgtcccaggaccttagaacctctctcaacaaccacg 323
DB 10328 TGGCGCGCTCGAGTCCGGCGCTCCGACCTCCAGACCTACTCTGCTCCCTGGGGCGACA 10387

QY 324 acgtcacgttgctgtctactccaccctcttgatggcgggcgagggtgcgtgatattact 383
DB 10388 GGCTCTCCACC GCCCGCCCTCAACAGCCCCA-----GGGTACCTCTGTTATCCGGCGAGC 10441

QY 384 ccacaggtccagagccctctctgggtcgagccccccacgtgctgcatattact 443
DB 10442 CCGCCGCGTGACGCGGTGCTCGACGTCTCACGCCACCAAGGTGTTGCGCCGCAAGA 10501

QY 444 ttacttcaacgggtacgacagagatcgacgcgggttcocgacgaaaatcgagctccctc 503
DB 10502 TCGCGGTGCAGTACGCCCTCCACCTCCGCCAGATGGACGCCGTCCAAGACGAGCTCGCG 10561

QY 504 ggcttcaccc 514
DB 10562 CAGGTCTAGCC 10572

RESULT 10
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligdon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation

```



```
Best Local Similarity 47.4%; Pred. No. 0.0078;
Matches 176; Conservative 0; Mismatches 189; Indels 6; Gaps 1;

Qy 144 ggccgcatggaacacacagcctccgcgtgcccgaacccacccggtcgtcgaattcg 203
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10208 GCCACAGCAGGCGAGATCGCGCGCGCTTCGTCGACAGGCGCTCTCTCCCTCGAGGACG 10267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 204 tggcgttctccacagctacgacgagcggtgaaagccctcgccgacgaggaagcgctaca 263
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10268 CGCGCGCATCGCGCGCTCGCAGAAAGCGCTCACACCGCTGGGGGCAACGGGGGCA 10327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 264 tgcagagatgaagccccgggctccgggcttaagaacccctcttctcaacaaccacg 323
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 TGCGCGCTCGAGCTCGCGCGCTCGACCTCCAGACCTACCTCGCTCCCTCGGGGGGCA 10387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 324 acgtcagttcgtctactccactcttctgcatggcgagggaggtggcggtgagt 383
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 GGCTCTCCACCGCGCTCACAGCCCA-----GGGTACCTCGTATCGGGGAGC 10441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 384 cccaggtcccgagcgccctctctggtgagcccgccacogtctgtgcatattact 443
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10442 CGCGCGCGCTCGACGCGCTCGACGCTCCTACCGCACCAAGGTGTTCGCCCGCAAGA 10501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 444 tctactcaacggctacgacgagatcgacgcggttccgacgaattcagctccctc 503
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10502 TCGCGGTGCGACTAGGCTCCGACTCCGCCAGATGGAGCGCTCCAAAGACGAGCTCGCG 10561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 504 ggcttccaccc 514
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10562 CAGGTCTAGCC 10572
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas M.
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458.076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-458-076A-6

Query Match 3.1%; Score 46.6; DB 2; Length 28958;
Best Local Similarity 47.4%; Pred. No. 0.0078;
Matches 176; Conservative 0; Mismatches 189; Indels 6; Gaps 1;

Qy 144 ggccgcatggaacacacagcctccgcgtgcccgaacccacccggtcgtcgaattcg 203
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10208 GCCACAGCAGGCGAGATCGCGCGCGCTTCGTCGACAGGCGCTCTCTCCCTCGAGGACG 10267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 204 tggcgttctccacagctacgacgagcggtgaaagccctcgccgacgaggaagcgctaca 263
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10268 CGCGCGCATCGCGCGCTCGCAGAAAGCGCTCACACCGCTGGGGGCAACGGGGGCA 10327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 264 tgcagagatgaagccccgggctccgggcttaagaacccctcttctcaacaaccacg 323
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 TGCGCGCTCGAGCTCGCGCGCTCGACCTCCAGACCTACCTCGCTCCCTCGGGGGGCA 10387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 324 acgtcagttcgtctactccactcttctgcatggcgagggaggtggcggtgagt 383
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 GGCTCTCCACCGCGCTCACAGCCCA-----GGGTACCTCGTATCGGGGAGC 10441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 384 cccaggtcccgagcgccctctctggtgagcccgccacogtctgtgcatattact 443
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10442 CGCGCGCGCTCGACGCGCTCGACGCTCCTACCGCACCAAGGTGTTCGCCCGCAAGA 10501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 444 tctactcaacggctacgacgagatcgacgcggttccgacgaattcagctccctc 503
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10502 TCGCGGTGCGACTAGGCTCCGACTCCGCCAGATGGAGCGCTCCAAAGACGAGCTCGCG 10561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 504 ggcttccaccc 514
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10562 CAGGTCTAGCC 10572
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


THIS PAGE BLANK (uspro)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2000, 21:13:28 ; Search time 3616.02 Seconds
(without alignments)
1837.825 Million cell updates/sec

Title: US-09-147-955-1
Perfect score: 1507
Sequence: 1 gaataattccacaaaatgg.....ttataatttagtaacaaaa 1507

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues 10495684
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

1:	em_est1:
2:	em_est2:
3:	em_est3:
4:	em_est4:
5:	em_est5:
6:	em_est6:
7:	em_est7:
8:	em_est8:
9:	em_est9:
10:	em_est10:
11:	em_est11:
12:	em_est12:
13:	em_est13:
14:	em_est14:
15:	em_est15:
16:	em_est16:
17:	em_est17:
18:	em_est18:
19:	em_est19:
20:	gb_est1:
21:	gb_est2:
22:	gb_est3:
23:	gb_est4:
24:	gb_est5:
25:	gb_est6:
26:	gb_est7:
27:	gb_est8:
28:	gb_est9:
29:	gb_est10:
30:	gb_est11:
31:	gb_est12:
32:	gb_est13:
33:	gb_est14:
34:	gb_est15:
35:	gb_est16:
36:	gb_est17:
37:	gb_est18:
38:	gb_est19:
39:	gb_est20:
40:	gb_est21:
41:	gb_est22:
42:	gb_est23:
43:	gb_est24:
44:	gb_est25:
45:	gb_est26:
46:	gb_est27:
47:	gb_est28:
48:	gb_est29:
49:	gb_est30:
50:	gb_est31:
51:	gb_est32:
52:	em_est20:
53:	em_est21:
54:	em_est22:
55:	em_est23:
56:	em_est24:
57:	em_est25:
58:	em_est26:
59:	gb_est33:
60:	gb_est34:
61:	gb_est35:
62:	gb_est36:
63:	gb_est37:
64:	gb_est38:
65:	em_est27:
66:	em_est28:
67:	em_est29:
68:	em_est30:
69:	gb_est39:
70:	gb_est40:
71:	gb_est41:
72:	gb_est42:
73:	gb_est43:
74:	gb_est44:
75:	em_est31:
76:	em_est32:
77:	em_est33:
78:	em_est34:
79:	gb_est45:
80:	gb_est46:
81:	gb_est47:
82:	em_est35:
83:	em_est36:
84:	em_est37:
85:	gb_est48:
86:	gb_est49:
87:	gb_est50:
88:	gb_est51:
89:	gb_est52:
90:	gb_est53:
91:	gb_est54:
92:	gb_est55:
93:	gb_gss1:
94:	gb_gss2:
95:	gb_gss3:
96:	gb_gss4:
97:	em_gss1:
98:	em_gss2:
99:	em_gss3:
100:	em_gss4:
101:	gb_gss5:
102:	gb_gss6:
103:	gb_gss7:
104:	gb_gss8:
105:	gb_gss9:
106:	em_gss5:
107:	em_gss6:
108:	em_gss7:
109:	em_gss8:
110:	em_gss9:
111:	em_gss10:
112:	em_gss11:
113:	gb_gss10:
114:	gb_gss11:
115:	em_gss12:
116:	gb_gss12:

```
117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description	
		Match	Count					
1	261.6	17.4	647	40	AI488782	EST247121	AI488782 EST247121	
2	237.2	15.7	536	44	AI771830	EST252930	AI771830 EST252930	
3	219.2	14.5	546	45	AI897621	EST367064	AI897621 EST367064	
C	4	214.2	14.2	747	71	AW349414	AW349414	AW349414
	5	211.8	14.1	559	45	AI896013	EST265456	AI896013 EST265456
6	209.2	13.9	478	45	AI899102	EST268545	AI899102 EST268545	
7	199.2	13.2	520	40	AI487571	EST445893	AI487571 EST445893	
8	194.4	12.9	649	72	AW442038	AW442038	AW442038	
9	191.2	12.7	500	40	AI489014	EST247353	AI489014 EST247353	
10	188.4	12.5	685	63	AW034683	AW034683	AW034683	
11	176.8	11.7	476	40	AI483541	EST278414	AI483541 EST278414	
12	167.4	11.1	517	44	AI779099	EST294930	AI779099 EST294930	
13	162.2	10.8	658	79	AW651280	EST295978	AW651280 EST295978	
14	150.8	10.0	612	63	AW034633	EST329734	AW034633 EST329734	
15	149.8	9.9	566	45	AI896026	EST278317	AI896026 EST278317	
16	149.4	9.9	602	63	AW031351	EST265469	AW031351 EST265469	
17	149.4	9.9	605	69	AW220656	EST274805	AW220656 EST274805	
18	149.2	9.9	518	63	AW032936	EST297125	AW032936 EST297125	
19	149.2	9.9	534	72	AW442216	AW442216	AW442216	
20	149.2	9.9	558	63	AW030266	EST373521	AW030266 EST373521	
21	149.2	9.9	576	63	AW031286	EST274561	AW031286 EST274561	
22	149.2	9.9	579	63	AW032414	EST275953	AW032414 EST275953	
23	149.2	9.9	606	63	AW035724	EST281878	AW035724 EST281878	
24	149.2	9.9	613	63	AW035364	EST280926	AW035364 EST280926	
25	149.2	9.9	642	79	AW651250	EST329704	AW651250 EST329704	
26	149.2	9.9	644	69	AW220874	EST297343	AW220874 EST297343	
27	147.6	9.8	591	79	AW648641	AW648641	AW648641	
28	147.6	9.8	596	64	AI775659	EST327191	AI775659 EST327191	
29	145.8	9.7	540	44	AI775659	EST299163	AI775659 EST299163	
30	145.6	9.7	518	63	AW030636	EST256759	AW030636 EST256759	
31	144.2	9.6	534	63	AW031171	EST273891	AW031171 EST273891	
32	141.2	9.4	680	63	AW030131	EST274709	AW030131 EST274709	
33	140.6	9.3	467	45	AI896353	EST273386	AI896353 EST273386	
34	139.8	9.3	548	44	AI778526	EST265796	AI778526 EST265796	
35	136.8	9.1	627	71	AW398421	EST259405	AW398421 EST259405	
C	36	136.2	9.0	619	72	AW459541	AW459541	AW459541
	37	134	8.9	513	45	AI895821	EST265264	AI895821 EST265264
38	134	8.9	519	44	AI772231	EST253331	AI772231 EST253331	
39	133.8	8.8	544	69	AW216889	AW216889	AW216889	
C	40	132.6	8.8	466	62	AW440325	AW440325	AW440325
C	41	132	8.8	537	79	AW650189	AW650189	AW650189
42	131	8.7	534	63	AW030442	AW030442	AW030442	
C	43	127.2	8.4	435	79	AW650188	AW650188	AW650188
C	44	125	8.3	523	47	AI997552	AI997552	AI997552
	45	124	8.2	287	64	AW127679	M110425	AW127679 M110425

ALIGNMENTS

RESULT	1
LOCUS	AI488782
DEFINITION	AI488782 647 bp mRNA EST247121 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
	EST 29-JUN-1999
	CLD18D16, mRNA sequence.

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 196 a 82 c 114 g 154 t
ORIGIN

Query Match 14.5%; Score 219.2; DB 45; Length 546;
Best Local Similarity 67.5%; Pred. No. 6.9e-51;
Matches 343; Conservative 0; Mismatches 153; Indels 12; Gaps 2;

```

Qy 579 tgaagagagcgtggaaactttgacggtgaaagagcgaagctgtgtgacacgt 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 50 TCAAGAGCAACTAGACACATAGATGAGTGGTGAAGAAATCCAAAGGTACTTGTGAATACAT 109

Qy 639 ttgatcgcttggagcccgatgacacgctattgataggatgagttgacgggacg 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 TTGATGCATTAGAGCTAGAGCCACTCAAGACTATTGAAATATACAAATTTAATTGCAATG 169

Qy 699 ggccttgattccctccgcttcttgacggtggagatccctccgaaacgctttcacggc 758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 GACCATTGATTCCTTCATCATCTTGGTGGAAGATTCATTCGAATCTTCATTTGGTG 229

Qy 759 gcatcttttcgaaatacggagagaaataactgcgtggagctgtgtgacacgagcga 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 GTGATCTTTTCAAAAGTCAAAAGT---ATGACTATCATGGAATGGTTAAACACAAAGCCTA 286

Qy 819 aatctcggtgtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 AATCATCAATTCCTTATATCTCATCTTTGGGAGCTATTCGAATTTATCAAGAAACCAAGG 346

Qy 879 aagagatgggaaaggcgtattagctgcggagagcgcgtgtgtgtgtgtgtgtgtgtgt 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 AGGAGATGGCAAAAGGGTGTATAGAGATCAAAAGGCCATTCATTATGGGTATATAAGAG--- 403

Qy 939 agaagaatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 -----ATCACGAAGAAGAAAGAAAGAAAGAAATTCAGTTGCATGATGAATAGAGA 457

Qy 999 aaatgggaaataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 AGCAAGGAAATAGTATACCATGTGTTCACACTGTAAGCTCTGCACATCCATCTTTAG 517

Qy 1059 gatgttcgtgacgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 GATGCTTGTCTGCACACTGTGGATGAA 545

```

RESULT 4
AW349414/c 747 bp mRNA EST 01-FEB-2000
LOCUS GM210007A20D2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
DEFINITION AW349414
ACCESSION AW349414
VERSION AW349414.1 GI:6847124
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 747)
AUTHORS Vodkin, L., Kelm, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V., Erpelting, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT On Jul 9, 1999 this sequence version replaced gi:5434916.
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomeystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

Location/Qualifiers source
1. 747
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3947"
/clone_lib="Gm-r1021"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Library Gm-r1021 is a sequence-driven, rerecked set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stragene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Kelm & Virginia H. Coryell, Department of Biology, Box 5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.kelm@uau.edu, virginia.coryell@uau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota, http://www.cbc.umn.edu/research/projects/soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html." BASE COUNT 206 a 235 c 97 g 187 t 22 others
ORIGIN

Query Match 14.2%; Score 214.2; DB 71; Length 747;
Best Local Similarity 59.7%; Pred. No. 1.9e-49;
Matches 414; Conservative 0; Mismatches 243; Indels 36; Gaps 3;
Qy 707 attccctccgcttcttggacggtggagatccctccgaaacgtcttaccggtgacatctt 766
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 ATTCNNNNNGCTTCNNNGCGTAAAGATCCGCGGATACCTNNNTATGCGGTGATNN 674
Qy 767 ttcgaaaaatcggaggagaataactgcgtggagtggtggacacgaaagccaaatcttcg 826
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 NNCNNTGCTTC-----AAATGATTACGTGTAATGGTGGACTACACCCCTGAGTTATCT 620
Qy 827 gtggtgatgtgctgt 886
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 GTGGTTATGTTTCATTTGGTACCTTGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 560
Qy 887 gggaaaaagggctattagctcgtggaaagggccgtttttatggatgatcagagacagaagaat 946
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 559 GCACGCGCTGCTAGATTCGGATATCTCTTCTTGTGGTGTATGAGATG----- 504
Qy 947 gacgacggcgaagaagaagaagagtggtgagttggttcattgggggaattgaaaaaattgggg 1006
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 -----CAAGCAATAGAAGATAACTGCAGAGAGGAAGTGCAGCAGAGGGGT 461
Qy 1007 aaatagttctgt 1066
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 AAGATTGTGAATGGT 401
Qy 1067 gtgacgcatgt 1126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 GTAACGCATTGTGGATGGAATTCGACTATGGAAGATTTGGGTTTCGGGGTTCCTATGCTG 341

```

QY 1127 ggggtgcccagtggtttgatcagacagcaaatcgaaagctgattgagagtcgtggggg 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 GCGTTTCGCGAGTGCAGACACCAAGGACGAATGCAAGATGCTGCGAAG 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1187 acaggggtgagagtgagaaagaa-----tgaaggggggTgggttgatggatctgag 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 AGGGGGTGAGGGTGGATGATTAAGGTGAATGTTGAGGAGGATAGTGAAGCAGAGGAG 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1238 atagagaggtgtgtggagatgggtgatggatgggggtgagaagacaaactagttagagaa 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 ATTAGGAAGTCTTTGGATGTTGGTTCATGGTGGAGAGTGGAGGAAAAGGACAGGAATTCAGAAAG 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1298 atgccaLaaaatgaaagactttggccagagaaagccatgggagagatggatcttcactc 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 AATGCTGATAAATGAAATGCTTGGCCAGGAAAGCCGTCACGGAAGGAGGCTCTTCGAT 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1358 aagaatctcaagcgttttcttctcatcaatggtgca 1390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 AGTAATATGAGGACTTTTCTCCATGATGTGCA 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
LOCUS AI896013 559 bp mRNA EST 27-JUL-1999
DEFINITION EST265456 tomato callus, TAMU Lycopersicon esculentum cDNA clone
ACCESSION CLC13H13, mRNA sequence.
VERSION AI896013
KEYWORDS AI896013.1 GI:5601915
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 559)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
Location/Qualifiers
1..559
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC13H13"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRP"
/notes="vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
201 a 84 c 112 g 162 t

```

BASE COUNT
ORIGIN

Query Match 14.1% Score 211.8; DB 45; Length 559;
Best Local Similarity 67.1%; Pred. No. 8.4e-49;
Matches 335; Conservative 0; Mismatches 152; Indels 12; Gaps 2;

```

QY 579 tgaaggaagaagctgaaacatttagacgggtgaagagaagcgaaagtgttggtgaacacgt 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 TCAAGAGACCAACTAGACACATTTAGATGGTGAAGAAATCCAAAGGTACTGTGTGAATACAT 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 639 ttgagctgttagagcccgatgcactcagctctattataggtataggtataggtataggtatag 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 TTGATGCATTAAGAGCTAGAGCCACTCAAGCTATTGAAAAATACAAATTTTAATTGAATTG 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 699 ggcctgttattccctccgctcttcttggacggcgagatccctccgaaacgtcttctacggcg 758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 GACCAATTGATTCCTTCATCATCTCTTGGTGTGAAAAGATTCAATTGGAATCTTCAATTGGTG 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 759 gcgatcttttcgaaacatcgagagagaataaactcgctggagtggttggacacgaagccga 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 GTGATCTTTTCAAAAGTCAAAATG---ATGACTACATGGAATGTTAAACACAAAGCCTA 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 819 aatcttcggtgtgtatgtcgttttggggagcgtttttggagggttttccaaaggcacaataag 878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 AATCATCAATTTGTTTATATCTCATATTTGGGAGTCTATTGAATTTATCAAGAAAACCAAAAGG 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 879 aagagattgggaaagggcctattagcctcggaagggcgctttttatgatgatgatacagaac 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 AGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATTCCTTATGGTAAAGAG--- 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 939 agaagaatgacgagcgcgagagaagaagaagaagattgagattgagattgagaaatgaaaa 998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 -----ATCAAGAAAGAAAAGAAAGAAAGAAATTAAGTTGCATGATGGAATTAGAGA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 999 aaatgggaaaaatagttcgtggtcgtcgagttggaggttctgcgcacccctgcgttgg 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 ACCAAGGGAATAGTACCATTGGTGTTCACAACTTGAAGTCTCGACACATCCATCTTTAG 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1059 gatgttcgtgacgattg 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GATGATTTGTCTCGACATG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
LOCUS AI899102 478 bp mRNA EST 27-JUL-1999
DEFINITION EST268545 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
ACCESSION CLC13FA16, mRNA sequence.
VERSION AI899102
KEYWORDS AI899102.1 GI:5605004
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 478)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Feb 18, 1999 this sequence version replaced gi:4296853.
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
Location/Qualifiers
1..478
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC13FA16"

FEATURES
source

```

```

/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT      172 a   57 c   102 g   137 t
ORIGIN

```

```

Query Match      13.9%; Score 209.2; DB 45; Length 478;
Best Local Similarity 67.3%; Pred. No. 4.3e-48;
Matches 330; Conservative 0; Mismatches 146; Indels 12; Gaps 2;

QY 606 gtgaagagagcgaaagttcttggtgaacacgtttgatcgcttgagcccgatgcactca 665
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GTGAAGAAAATCCAAAGGTACTTGTGAATACATTTTGATGCATTAGAGCTAGAGCCACTCA 60

QY 666 cggctattgatagatagattgatcggtcggtcggtgattccctccctcttgg 725
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 AAGCTATTGAAAATACAAATTAATTGGAATTGGACCATGATTCTTCATCATCTCTGG 120

QY 726 scggcgagatccctcgaaacgtcttaccggcggtgatctttcgaaaaatcgagagaga 785
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GTGGAAGATTCATTGGAATCTTCATTTGGTGGTATCTTTTTCAAAAGTCAAAATG--- 177

QY 786 ataactcgtgagtggttgacacgaagccgaattctcgtggtgattgctgttg 845
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 ATGACTACATGTAATGTTAAACACAAAGCCTAAATCATCAATTTCTTATATCTCATTTG 237

QY 846 ggagcgtttgaggtttccaaagcccaatggagagattgggaaaggggtatagcct 905
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 GGAGTCTATTGAATTTATCAAGAAACCAAAAGGAGAGATTGCAAAAGGGTGTAGAGA 297

QY 906 ggggaagcgctttttatgatgatcagagacagagaagatgacgcgcgaagaagaag 965
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 TCCAAAGGCCATCTTATGTGTAATAAGAG-----ATCAAGAAGCAAGAAAAGAG 348

QY 966 aagaagattgattgattggttggaattgaaataatggggaataatgcttgcgtgct 1025
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AAGAAATTAAGTTGATGATGGAATTAGAGAAGCAAGGAAATAGTACCATGGTGT 408

QY 1026 cgcagttggaggttctgcgcacccctgcgttggtgatttctgcgcatgtggtgga 1085
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 CACAACCTGAAGTCTGCACACATCCATCTTTAAGATGTTTCTCTCGCACTGTGATGGA 468

QY 1086 actcggctgt 1095
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 ATTGCAGCTCT 478

```

```

RESULT 7
LOCUS      A1487571
DEFINITION EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            CLED10D21, mRNA sequence.
ACCESSION  A1487571
VERSION    A1487571.1 GI:4382942
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
            Solanaceae; Solanum; Lycopersicon.
REFERENCE  1 (bases 1 to 520)
AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Holt,I.E., Liang,F., Upson,J., Ronning,C.M., Craven,M.B.,
            Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
            Martin,G.B., Tanksley,S.D. and Giovannoni,J.
            Generation of ESTs from tomato carpel tissue
TITLE

```

JOURNAL COMMENT

Unpublished (1999)
On Apr 7, 1998 this sequence version replaced gi:3035667.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4393
Email: dfrisch@CLEMSON.EDU.

FEATURES

Location/Qualifiers
1..520

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED10D21"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 194 a 76 c 107 g 143 t
ORIGIN

Query Match 13.2%; Score 199.2; DB 40; Length 520;
Best Local Similarity 67.2%; Pred. No. 2.9e-45;
Matches 317; Conservative 0; Mismatches 143; Indels 12; Gaps 2;

```

QY 579 tgaaggagagcgtaaacctttagacggtgagagaaagcgaaagtggtggaacagt 638
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 TCAAGAGCAACTAGACACATTAGATGGTGAAGAAAATCCAAAGGTACTTGTGAATACAT 116

QY 639 ttgatcggttgagcccgatgcactcacgcttattgataggtagtgatcgggatcg 698
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 TTGATGCATTAGAGCTAGAGCCACTCAAGACTATTGAAAAATACAAATTAATTGGAATTG 176

QY 699 ggcgttgattccctcgcttattgagcggtgagagatccctccgaaacgtcttaacgcg 758
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 GACCATTGATTCCTTCATCATCTTGGGTGAAAAGATTCATTGGAATCTTCAATTTGGTG 236

QY 759 gcgatcttttcgaaaaatcgaggagaaataaactgcgtggtggtgacacgaagccga 818
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GTGATCTTTTTCAAAAGTCAAAATG---ATGACTACATGGAATGGTTAAACACAAAGCCTA 293

QY 819 aatctcgggtggtatgtctgtttggagcggttttgaggtttccaaagcacaacatcg 878
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 AATCATCAATTGTTTATCTCATCTTGGGAGCTATTGAAATTTATCAGAAACCCAAAGG 353

QY 879 aagagattgggaaggggtattagcctgcggaaggcgttttttttattgatatcagagaac 938
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 AGGAGATTGCAAAAGGGTTGATAGATCCAAAGGCCATCTTATGGGTAATAGAG--- 410

QY 939 agaagaatgacgcgcgaagaagaagaagaagatgtagtgcattggggaattgaaaa 998
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 -----ATCAGAGAGAGAAAAGAGAGAGAGAAATTAAGTTGCATGATGATGATAGAGA 464

QY 999 aaatgggaaaaatagttctggtctcagttgaggttgcagttggtcgcaccc 1050
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 AGCAAGGGAAAAATAGTACCATTGTTTCAACTTGAAGTCTGCACATCC 516

```

RESULT 8
LOCUS AW442098
DEFINITION EST311494 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
 clone CLEN21G12 5', mRNA sequence.
ACCESSION AW442098
VERSION AW442098.1 GI:6977349
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 649)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5407380.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.

FEATURES
source
1. 649
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEN21G12"
/clone_lib="tomato fruit red ripe, TAMU"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Giovannoni. Fruit were tagged at the breaker stage (first sign of lycopen accumulation on the blossom end of the fruit) and harvested 7 days post-breaker (fully red-ripe), 10 days post breaker, and 20 days post-breaker (over-ripe). 20 day fruit which showed external or internal signs of pathogenesis were discarded. Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT 217 a 118 c 124 g 190 t
ORIGIN

Query Match 12.9%; Score 194.4; DB 72; Length 649;
Best Local Similarity 60.6%; Pred. No. 6.8e-44;
Matches 390; Conservative 0; Mismatches 221; Indels 33; Gaps 3;

Qy 326 gtacgttctgtctactccacacctttgcatggcgccgaggtggcggtgagtc 385
Db 1 GTCACGTCCCTCGCTACACTCTTCTGCTACTTGGGAGGTAGCGGTGAACTC 60

Qy 386 cagggtcccgagcccttctgtgtgagccgccacccgtgtgtgtcatattacttc 445
Db 61 CATATCCCATCCGCGTACTATGGATTCAACGCAACTGTGCTAGACATATACTAT 120

Qy 446 tacttcaacggtcacgacagagatcgacgcccgtttccgacga-----a 490
Db 121 TACTTCAATGGCTATGAGATGAATGAAGTGTAGTAGCTCAATGATCCAAATGGAGT 180

Qy 491 attcagctccctcggttccacctggagcagcgcagcttcttcgaaccttctgtgcgcg 550
Db 181 ATCCAATTGCCAAGGCTCCATTACTTAAAGCAAGCATCTTCCATCTTTTGTAGTTCA 240

Qy 551 gagacccggagagattccggttgatgatg-----aagagagacgtggaa 595
Db 241 TCTAGCTCAAAAGATGATGAATGATGATTTGCTTACCACATTCAAAGAGCAACTAGAC 300

Qy 596 accttagacggtgaagagagcgaaagtgttggtagaacggtttagcgttggagccc 655
Db 301 ACATTAGATGTTGAAGAAAATCCAAAGGTTACTTGTGAATACATTTGATGCTTAGAGCTA 360

Qy 656 gatgcactcagcgtattatgaggtatgattgattcgaggtcgccgttgatccctcc 715
Db 57 TCAAGAGCAACTAGACATTTAGATGGTGAAGAAAATCCAAAGGTTACTTGTGTATACAT 116

Db 361 GAGCCACTCAAGGCTATTGAAAAATACAAATTAATTGGAATTGGACCATTGATTCCTTCA 420

Qy 716 gccttttggacggcgagatccctccgaacacgttcttacggcgcgatcttttcgaaaaa 775
Db 421 TCATTCTTGGGTGGAAGAGATTCAATGGAAATCTTCTATTTGGTGTGATCTTTTTCAAAAG 480

Qy 776 tcggagagaaataactgcgtgagtggttgacacgaagccgaatacttcggtgggtgat 835
Db 481 TCAATG--ATGACTACATGGAATGGTTAAACACAAAGCCTAATCATCAATGTTAT 537

Qy 836 gtgtcgtttggagcgttttggaggttccaaaggccacaaatggagsgattggaaaggg 895
Db 538 ATCTCATTTGGGAGTCTATTGAAATTTATCAAGAAACCAAGAGGAGATTGCAAAAGG 597

Qy 896 ctattagcctgcggaagccgcttttttgatggatcacgagaaca 939
Db 598 TTGATAGAGATCCAAAGGCCATTCTTATGGGTAAATAGAGATCA 641

RESULT 9
AI489014
LOCUS EST247353 tomato ovary, TAMU Lycopersicon esculentum cDNA clone 29-JUN-1999
DEFINITION CLED19P7, mRNA sequence.
ACCESSION AI489014
VERSION AI489014.1 GI:4384385
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 500)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Viston,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Niemman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT On Nov 2, 1998 this sequence version replaced gi:3830982.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.

FEATURES
source
1. 500
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED19P7"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cLED - tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 188 a 69 c 105 g 138 t
ORIGIN

Query Match 12.7%; Score 191.2; DB 40; Length 500;
Best Local Similarity 67.1%; Pred. No. 5e-43;
Matches 306; Conservative 0; Mismatches 138; Indels 12; Gaps 2;

Qy 579 tgaagagagcgtggaaacttttagcgttgaagagagcgaaagcgaagctgttgtaaacgt 638
Db 57 TCAAGAGCAACTAGACATTTAGATGGTGAAGAAAATCCAAAGGTTACTTGTGTATACAT 116

[illegible]

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 1, 2000, 21:31:42 ; Search time 61.65 Seconds
(without alignments)
526.332 Million cell updates/sec

Title: US-09-147-955-12
Perfect score: 2475
Sequence: 1 MVQPHVILTFPAQGHINFA.....EGSSHNKRAFDVAKGF 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.organism:*
10: sp.plant:*
11: sp.protist:*
12: sp.virus:*
13: sp.unclassified:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1363.5	55.1	460	10 Q92R27	Q92R27 perilla fru
2	1320	53.3	461	10 Q92R25	Q92R25 verbena x h
3	1307.5	52.8	443	10 Q92R26	Q92R26 perilla fru
4	1207	48.8	467	10 P93709	P93709 nicotiana t
5	1133.5	45.8	474	10 O04930	O04930 arabidopsis
6	1090.5	44.1	456	10 O23270	O23270 arabidopsis
7	1084.5	43.8	519	10 Q92VY2	Q92VY2 arabidopsis
8	1060	42.8	455	10 Q92VY5	Q92VY5 arabidopsis
9	1049.5	42.4	458	10 O23406	O23406 arabidopsis
10	809.5	32.7	479	10 O23401	O23401 arabidopsis
11	777	31.4	449	10 O22822	O22822 arabidopsis
12	766	30.9	475	10 O23402	O23402 arabidopsis
13	762	30.8	460	10 O48676	O48676 arabidopsis
14	761	30.7	484	10 O23400	O23400 arabidopsis
15	733	29.6	449	10 O22820	O22820 arabidopsis
16	727	29.4	456	10 O22182	O22182 arabidopsis
17	712	28.8	438	10 O22183	O22183 arabidopsis
18	645	26.1	453	10 O22186	O22186 arabidopsis
19	592	23.9	481	10 Q92WJ3	Q92WJ3 arabidopsis

20	553.5	22.4	455	10	O64733	O64733 arabidopsis
21	542	21.9	482	10	Q92UV0	Q92UV0 arabidopsis
22	530.5	21.4	496	10	Q92Q96	Q92Q96 arabidopsis
23	525	21.2	440	10	O64732	O64732 arabidopsis
24	521	21.1	495	10	Q92Q94	Q92Q94 arabidopsis
25	513	20.7	491	10	Q92Q99	Q92Q99 arabidopsis
26	512.5	20.7	495	10	Q92Q95	Q92Q95 arabidopsis
27	508.5	20.5	496	10	Q92Q97	Q92Q97 arabidopsis
28	502.5	20.3	454	10	Q9XF16	Q9XF16 forsythia x
29	500	20.2	462	10	O04622	O04622 arabidopsis
30	498	20.1	452	10	O22303	O22303 vitis vinif
31	497	20.1	456	10	O22304	O22304 vitis vinif
32	496	20.0	476	10	P93364	P93364 nicotiana t
33	492.5	19.9	496	10	Q92Q98	Q92Q98 arabidopsis
34	491	19.8	476	10	P93365	P93365 nicotiana t
35	483.5	19.5	484	10	Q92Q94	Q92Q94 arabidopsis
36	483	19.5	477	10	Q92WQ5	Q92WQ5 vigna mungo
37	482	19.5	483	10	Q9XES4	Q9XES4 malus domes
38	477	19.3	466	10	O43526	O43526 lycopersico
39	457	18.5	455	10	Q92WS2	Q92WS2 vigna mungo
40	443.5	17.9	488	10	P93789	P93789 solanum tub
41	443	17.9	460	10	Q92Q93	Q92Q93 arabidopsis
42	432	17.5	381	10	Q92WQ3	Q92WQ3 vigna mungo
43	427.5	17.3	474	10	O82382	O82382 arabidopsis
44	427	17.3	478	10	O49492	O49492 arabidopsis
45	425	17.2	447	10	O04114	O04114 perilla fru

ALIGNMENTS

RESULT 1
Q92R27 PRELIMINARY; PRT; 460 AA.
ID Q92R27
AC Q92R27
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE.
GN PF3R4.
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. SHIKUN; TISSUE=LEAF;
RX MEDLINE; 99167509;
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin.";
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013596; BAA36421.1;
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 460 AA; 50974 MW; FE7CCF22 CRC32;

Query Match 55.1%; Score 1363.5; DB 10; Length 460;
Best Local Similarity 57.8%; Pred. No. 1.5e-91;
Matches 273; Conservative 74; Mismatches 106; Indels 19; Gaps 8;

Qy	1	MVQPHVILTFPAQGHINPALQFAKLVKNGIEVTFSTSIYAQSRMDEKSLNNA--PKGL 58
Db	1	MVRRVLLATFPAQGHINPALQFAKLLKGLDVTFTFSVYAWRRMANTASAGNPPGL 60
Qy	59	NFIPIFSGDFGDFHSDKDPVFMVSQLRKGGSETVKKILITCSENGQITCILYSIFLPWA 118
Db	61	DFVAFSDGYDDGLKPCGDKRYMSEMKARGSEALRNLL---NNHDVTFWYSHLPWA 116
Qy	119	AEVAREVHIPALLWSQPAITLDIYFFNFHGYEKAMNESNDPNWSQLPGLPLETDL 178

QY	235	YGIGPLIPSAFLGNDPLDASFGCDLFO---NSNDYMEWLANSKNSSVYVYISFGLMNP	291
Db	229	IAIGPLIPSAFLGDKDPSDRSFGCDLPEKGSNDDDCLEWLSNTNPRSSVYVYISFGSFYNT	288
QY	292	ISQMBEISKGLTIDGRPLFWIKENEGKEEENKKGCTEELEKIGKIVPMCSOLEVLKH	351
Db	289	KSQMBEIRAGLLDGRPLWVVRVN---EGEEVLISCMEELEKRVGKIVSWCSOLEVYLTH	344
QY	352	PSLGCFVSHCGWNSALSLSLGGVVPVAFPOWTDQMTNKAQVEDVWKSGVRVINEDG-VV	410
Db	345	PSLGCFVTHCGWNSTLESISFGVPMVAFPOWFDOGTNAKLMDYVRTGVRVANEEGSVV	404
QY	411	ESEETKRCIELVMDGGEKGBELRKNAKKWELAREAVKEGGSSHKNLKAFDDV	464
Db	405	DGDETRRTCTEEVMDGGSKRKLRESAGKWKDLARKAMEDEGSSVNNLKVFLDEY	458

RESULT	3
Q9ZR26	
ID	Q9ZR26
AC	PRELIMINARY; PRT; 443 AA.
DT	01-MAY-1999 (TRENBLrel. 10, Created)
DT	01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT	01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE	UDP-GLUCOSE:ANTHOCTANIN 5-O-GLUCOSYLTRANSFERASE HOMOLOGUE.
GN	PF3R6.

01- MAY-1999 (11EMBREL. 10, Last sequence update)
01- NOV-1999 (11EMBREL. 12, Last annotation update)
UDP-GLUCOSE:ANTHOCYANIN 5-O-GLUCOSYLTRANSFERASE HOMOLOGUE.
PF3R6.
GN
Perilla frutescens.
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SHIKUN; TISSUE-LEAF;
RX MEDLINE; 99167509.
RA YAMAZAKI M., GONG Z., FOKUCHI-MIZUTANI M., FUKUI Y., TANAKA Y.,
RA KUSUMI T., SAITO K.;
RT "Molecular cloning and biochemical characterization of a novel
RT anthocyanin 5-O-glucosyltransferase by mRNA differential display for
RT plant forms regarding anthocyanin."
RL J. Biol. Chem. 274:7405-7411(1999).
DR EMBL; AB013597; BAA36422.1; -
DR PROSITE; PS00375; UDPGT; 1.
SQ Transferase.
KW SEQUENCE 443 AA; 49110 MW; D87B5486 CRC32;

Qy	59	NF1PSDGFDEGFDHSDKDPVPMYSQLRKCSQETVKKIIILUTCSENGQPITCIIKISIFILPWA	118
Db	61	DFVARSQYDGLKPGGDKRYKSEMKAQSEALRMILL-----NNDVDVTFVYVSHLFAWA	116
Qy	119	AEVAREVHIPSALWQSOPATLIIYFNFGYEKANANESDPMWSIQLPGLPLETRDL	178
Db	117	AEVARSHPVTAIIWVEPATVLCIYHFYFNFYADETDAGSNE-----IQLPRLPSLEQSRSL	172
Qy	179	PSFLPLPYGAKGSLRVALPP-----FKELIDTLDAETTPKRLVNTFDELEPEALNAIEGYK	233
Db	173	PTFLPLP-----ATPERFLRMKEKLETLGCEKARKVLNTFDEALPDALTAIDRYE	223
Qy	234	FYGIQPLIPSAFLGNDPDLASFGCDLFQNS--NDYMEWLNSKPMSSVVYVYISFGSLMNP	291
Db	224	LIGIGPLIPSAFLGDEPDSYTSYGGDLFEKSENNCNCVWLNSKPMSSVVYVYISFGSVLRFP	293
Qy	292	ISOMEIEISKGLIDIGRPLPWYIK--NEKG--KEEENKLGCTEELEKIGKIVPWCISOLEV	348


```

Db 284 KAQMEIEIGKGLACGRFLMIRQKNDKDEEEELSCIGELKMKGIYWCQSLEY 343
QY 349 LKHPSLGCFVSHCGNNSALESACGVVPVAFPOWTDQMTNAKOVEDVWKGVRVRINEDG 408
Db 344 LAHPALGCFVTHCGNNSAVESLSCGIPVAVPQWFDQTTNAKLIEDAWGTGVRVRNEDG 403
QY 409 VYSEIEIKRCIELVMDGGEELRKNNAKWKELAREAV 447
Db 404 GVDGCEIERCVWMDGDKTKLVRENAIKWTLARQAM 442

RESULT 4
ID P93709 PRELIMINARY; PRT; 467 AA.
AC P93709;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE UDP-GLUCOSE:INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
GN JIGT.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BY-2; TISSUE-SUSPENSION CULTURE;
RA KOJIMA H., HASHIZUME K., IMANISHI S., NAKAMURA K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000623; BAA19155.1; -.
DR MENDEL; 9421; Nicta; 1525; 9421.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 467 AA; 52989 MW; 3726B31 CRC32;

Query Match 48.8%; Score 1207; DB 10; Length 467;
Best Local Similarity 50.1%; Pred. No. 3.7e-80;
Matches 234; Conservative 87; Mismatches 132; Indels 14; Gaps 7;

QY 5 HVILTFPAQHINPALQFAKNLVK--MGIEVTFSTSIYAQSRMDEKSIILNAPK--GLNTP 62
Db 9 HVILTFPAQHINPALQFAKNLVK--MGIEVTFSTSIYAQSRMDEKSIILNAPK--GLNTP 63
QY 63 FSDGDFEGFHSKDPV--PYNOLAKKCGSEVVKIILTCSENGOPIICLLYSIFLPAWAEV 121
Db 64 FSDGDFGNGKSGFDYHLFNSAISKHSEFANLIKSAKAGYPPFRVIYITLMDWAGSV 123
QY 122 AREVHPSALUWSQPATILDIYVFNHGYEKAMAN--ESNDPNWSIOLPGLPLETRDPLS 180
Db 124 AKKLHPSLTFWQIOPAFDIYVYRFTNFANYKNDSDQD--IIEPLGLSLSSSDPFS 181
QY 181 FLPLPGAGKSLRVLPPEKELIDTLDAETTPKILVNTFDELEPEALNATEGYKFGIGPL 240
Db 182 FV--FDVYKSNDAVESIKRQIEILNSEENPRILVNTFDELEPEALNATEGYKFGIGPL 239
QY 241 IPSAFGLNDPLDASFGDLFQNSNDYMWLNKSPNSVYVTSFGSLMNPSSIQMEESK 300
Db 240 IPSFLDKDRDNFADADMTESNTNEMWLDARANKSVIYAFGSYABISSQWMEESQ 299
QY 301 GLIDIGRPFLWIKENEKGEENKKGTELEKIGKVPWCQSLEYLKHPSLGCIFYSH 360
Db 300 GLKLCGRFLWVIRETLNGEKE--EKLTKDELEKIGRIVRWCSQMEVLKHSVGCPLTH 358
QY 361 CGWNSALESACGVVPVAFPOWTDQMTNAKOVEDVWKGVRVRINEDGVYSEIEIKRCIE 420
Db 359 CGWNSLTSLASGVPVACPIWPDQICNAKLIDQWIKIGVRVNAKKEGIIKRDFOKTE 418
QY 421 LVMDGGEELRKNNAKWKELAREAVKGGSHKMLKAFIDVAKG 467
Db 419 IVYGDAGEEELRKNNAKWKELAREAVKGGSHKMLKAFIDVAKG 465

```

```

RESULT 5
ID O04930 PRELIMINARY; PRT; 474 AA.
AC O04930;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE UDP-GLUCOSE:INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE
GN (EC 2.4.1.121) (INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE).
DE IAGLU.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA GILAD A., SCOLNIK P.A., BAR-ZVI D.;
RL Plant Physiol. 113:1004-1004(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA THORNBURG R.W., GRAHAM R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOLE-3-ACETATE -> UDP +
INDOLE-3-ACETYL-BETA-1-D-GLUCOSE.
DR EMBL; U81293; AAB58497.1; -.
DR MENDEL; 16583; Arabid; 2543; 16583.
DR PROSITE; PS00375; UDPGT; 1.
DR PFAM; PF00201; UDPGT; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 474 AA; 53740 MW; 35521D70 CRC32;

Query Match 45.8%; Score 1133.5; DB 10; Length 474;
Best Local Similarity 49.0%; Pred. No. 8.6e-75;
Matches 238; Conservative 77; Mismatches 120; Indels 51; Gaps 14;

QY 4 PHVILTFPAQHINPALQFAKNLVK--MGIEVTFSTSIYAQSRMDEKSIILNAPKGLNI 61
Db 12 PHVILTFPAQHINPALQFAKNLVK--MGIEVTFSTSIYAQSRMDEKSIILNAPKGLNI 70
QY 62 FSDGDFEGFHSKDPV--PYNOLAKKCGSEVVKIILTCSENGOPIICLLYSIFLPAWAEV 113
Db 71 FSDGDFEGFHSKDPV--PYNOLAKKCGSEVVKIILTCSENGOPIICLLYSIFLPAWAEV 130
QY 114 FLPLPGAGKSLRVLPPEKELIDTLDAETTPKILVNTFDELEPEALNATEGYKFGIGPL 173
Db 131 LUTWALAREFHLPSALLWQVPTVTSIFTHYENGEDALISEMANTPSSIKLPSLPL 190
QY 174 ETRDLPFLPYGAKGSLRVLPPEKELIDTLDAETTPKILVNTFDELEPEALNATEGY 232
Db 191 TVRDIPSFTVSSNVYAF--LPAPREQIDSLKEEINPKILINTFOELEPEANSSVDPNF 247
QY 233 KFGVIGLPSAFGLNDPLDASFGDLFQNSNDYMWLNKSPNSVYVTSFGSLMNPSSIQ 292
Db 248 KIVPGVPLU-----TLRTD-FSSRGYIEWLTADKADSSVLYVSGTFLAVLSK 293
QY 293 SQMEETSKGLIDIGRPFLWV-----KENEKGEENKKGTELEK-----IKGIYV 341
Db 294 KQIVELKALIQSRPFLWITDKSRNKEDEQEKED-----CISSEKSFDEIGMYSV 348
QY 342 WCSOLEVLKHPSLGCFVSHCGNNSALESACGVVPVAFPOWTDQMTNAKOVEDVWKGVR 401
Db 349 WCDQFRLNHRISGCFVTHCGNNSLTSLASGVPVAFPOWTDQMTNAKLIEDAWGTGVR 408
QY 402 V--RINEDG--VYSEIEIKRCIELVMDGGEELRKNNAKWKELAREAVKGGSHKML 457
Db 409 VMEKKEEGGVVYVDEEIRRCIEEYME--DKAEFRGNATRWKDLAAEAVREGSGSFNHL 466
QY 458 KAFIDD 463
Db 458 KAFIDD 463

```

Db 467 KAFVDE 472

RESULT 6
O23270

ID O23270 PRELIMINARY; PRT; 456 AA.

AC O23270;

DT 01-JAN-1998 (T-EMBLrel. 05, Created)

DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)

DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)

DE GLUCOSYLTRANSFERASE.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

OC Arabidopsis.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 98121113.

RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,

RA BERGMANN R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,

RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,

RA WEDLER E., WABUTT R., WEITZGESSER T., POHL T.M., TERRY N.,

RA GIELEN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,

RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HENDEL S., KOTTER P.,

RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,

RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,

RA VOUKELATOU E., MILIONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,

RA HILBERT H., DUESTERHOFF A., MOORES T., JONES J.D.G., ENEVA T.,

RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,

RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,

RA SCHUELLER C., CHALWATZIS N.

RT *Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of

RT Arabidopsis thaliana.*

RL Nature 391:485-488(1998).

RN [2]

RP SEQUENCE FROM N.A.

RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z97335; CAB10189.1; .

DR PROSITE; PS00375; UDCPT; 1.

DR PFAM; PF00201; UDCPT; 1.

KW Transferase.

SQ SEQUENCE 456 AA; 50702 MW; EAE0D380 CRC32;

Query Match

Best Local Similarity 44.1%; Score 1090.5; DB 10; Length 456;

Matches 221; Conservative 85; Mismatches 132; Indels 29; Gaps 9;

Qy 3 QPHVLTFFAAGHINPALQFAKNLKGIEVTFSTSYAOSRDEKILNAPGLNIP 62

Db 11 RPHVLTFFAAGHINPALQANLRLTHGATVYTAVSARHMRGEP-...STKGLSFAW 67

Qy 63 FSDGFDGDFHSDPVPYMSQLKCGSEVTKII--LTCSENGOPICLLYSIFLPWAA 119

Db 68 FTDGFDGLKSFEDQKIYMSQLKCGSEVTKII--LTCSENGOPICLLYSIFLPWAA 127

Qy 120 EVAREVHPSALLWSQATPILDIYFNFGYKEMANESNDPNNSIQLPLETDLPL 179

Db 128 TVAREFHPLPTLLWIEPATVLDIYFYFNFGYKEMANESNDPNNSIQLPLETDLPL 182

Qy 180 SFLPYGAKSLRVALPFPFKELDTDAETTPKILVNTFDELEPEALNAIEGYGIGP 239

Db 183 SFLQPSKA---LPSALVTLREHLEALETESNPKILVNTFSALEHDAELTSVEKLMIGP 239

Qy 240 LIPSAFLGGNDPLDASFGDGLFQNSN-DYMEWLNKSPNSVYVIFSGSLMN--PSISOME 296

Db 240 LVSS-----SEKTDLDFKSSDEDTKWLDKSLERSYIYISLGFHADDLPE-KHME 288

Qy 297 EISKGLIDIGRPFLLWIKENKCKEDENKGLGICIELEKIGKIVPWCQSOLBVLKHPSLGC 356

Db 289 ALTHGVLTATNPFLWIVREKNPEKKNRFLIRGSDR-GLVVGWCSQTAVLARCAGVC 347

Qy 357 FVSHCGWNSALESACGVPPVAPPQWTDQMTNAKQVEDVWKSQVVRINEDGVVSESEIK 416
Db 348 FVTHCGWNSLESLESGVPVAPPQADQCTTAKLVEDTWIRGVKVKVGEEDGVDGEIR 407

Qy 417 RCIELVMDGEGKEELKNAKWEAREVKEGGSSHKNLKAFIDD 463

Db 408 RCLEKVMSSGGEAEEMRENAEKWAKAVDAAEAGGSPDLNLKGFVDE 454

RESULT 7

Q92VY2

ID Q92VY2 PRELIMINARY; PRT; 519 AA.

AC Q92VY2;

DT 01-MAY-1999 (T-EMBLrel. 10, Created)

DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)

DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)

DE T25N20.20.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

OC Arabidopsis.

RN [1]

RP SEQUENCE FROM N.A.

RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,

RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,

RA VYSOTSKAYA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,

RA ECKER J.R.;

RT *Genomic sequence for Arabidopsis thaliana BAC T25N20.*;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC005106; AAC80599.1; .

DR PROSITE; PS00375; UDCPT; 1.

SQ SEQUENCE 519 AA; 58584 MW; C982007F CRC32;

Query Match

Best Local Similarity 43.8%; Score 1084.5; DB 10; Length 519;

Matches 232; Conservative 71; Mismatches 131; Indels 47; Gaps 14;

Qy 1 MVQPHVLTFFAAGHINPALQFAKNLKGIEVTFST--SIYASQSRMDEKILNAPKG 57

Db 51 MAPPHFLVTFPAAGHVNPSLRPARLIRTGARTVFTCVSVFHNMIANH--NKNVN 107

Qy 58 LNFIFPSDGFDEG-----FDHSDKDPVYMSQLKCGSEVTKIIITCSENGOPITCLYS 112

Db 108 LSFLTFSDGFDGGSISTYEDRQKRSV-----NLKVNDRKALSDFEATKNGDSPVTCIYT 163

Qy 113 IFLPWAEVAREVHPSALLWSQATPILDIYFNFGYKEMANESNDPNNSIQLPGL 172

Db 164 ILLNWAQVARRRQPSALLWIQIPALVFIYTHF-----MGNKS-----VFELPNS 212

Qy 173 LETRDLPSFLPYGA-KGSLRVALPFPFKELDTDAETTPKILVNTFDELEPEALNAIEG 231

Db 213 LEIRDLPFLTPSNTKG---AYDAFQEMERFLIKETPKILINTFDSLEPEALTAPN 268

Qy 232 YKFGICGLPISAFGLGNDPLDASFGDGLFQNSNDYMEWLNKSPNSVYVIFSGSLNPS 291

Db 269 IDWAVAGPLPTEIFS-----STNKSVDQSSSTTLMDSKTSSEVYVIFSGFWELS 322

Qy 292 ISOMEELSKGLIDIGRPFLLWV--KENEKKG---EEE---NKKLGCIELEKIGKIVPWC 343

Db 323 KQTEELARALIEGRPFLLWVTDKSNRETKEGEETEIEKTAGRHELEEVGMVSWC 382

Qy 344 SQLEVLKHPSLGCFVSHCGWNSALESACGVPPVAPPQWTDQMTNAKQVEDVWKSQVVR 403

Db 383 SQIEVLSHRVAGCFVTHCGWNSLESVLTGVPVAPPQWTDQMTNAKLEESWKTGVRR 442

Qy 404 INEDGVVSESEIKRCEIVMDGGERGEELKNAKWEAREVKEGGSSHKNLKAFIDD 463

Db 443 ENKDGVLVERGEIRRCLEAVME--EKSVELRENKAKWKLAMEAGREGGSSDKNMEAFVD 500

Qy 464 V 464

```

Db 501 I 501
RESULT 8
Q92VY5 PRELIMINARY; PRT; 455 AA.
ID Q92VY5
AC Q92VY5;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE T25N20.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC eumollusca; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,
RA VYSONSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T25N20.17";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005106; AAC80596.1; -.
DR PROSITE; PS00375; UDEGT; 1.
SQ SEQUENCE 455 AA; 51190 MW; F9BAC33E CRC32;

Query Match 42.8%; Score 1060; DB 10; Length 455;
Best Local Similarity 46.9%; Pred. No. 1.8e-69;
Matches 223; Conservative 78; Mismatches 136; Indels 38; Gaps 12;

Qy 1 MVQPHVLTTPAQGHINPALQFAKNLVK--MGIEVTFSTSIYAQSRMDEKSIILNAPKG 57
Db 1 MAOPHVLVTFPAQGHINPSLEAKRLAGTISGARVTFFAAISAYNR--MFSTENVETLIFA 70

Qy 58 LNFIPFSDGDEGFDHSDKPV--FYMSQLKCGSETVKKIILTCEGNGPITCLLYSIFLP 116
Db 58 LSFUTSDGDDGVISNTDQVQNLVHVRNGDKALSDFIENQNGSDSPVSCLIYILPN 117

Qy 117 WAAEVAEREVHIPALLWSQATILDIYFNFHGYEKAMANESNDPNMSIQLPGLPLETR 176
Db 117 WPKVARRHLPVHLWIOPAFAFDIY-----NYSTGNNSVFFPNLPSLEIR 166

Qy 177 DLPSFLPYGAGSLVALPPEKELIDTDAETTPKILVNTFDELEPEALNATIEGKFKYG 236
Db 167 DLPSFLP---SNTNKAQNVQELMDFLKEENPKILVNTFDSLEPEELTAIPNIEMVA 223

Qy 237 IGPLIPSAFAGNDPDLDAFGDGLFQN--SNDYMEWLNKSPNSVYIYFSGLSMNPISQ 294
Db 224 VGPLPAIEFTGES-----GDLSDHQSSSYTLWLDKSTESSYIVVSGTAVLSKKQ 278

Qy 295 MEETSKGLIDIGRPFLWVI--KENEKGK---EEE---NKLKGLCTEELEKIGKIVPWCQSOL 346
Db 279 IEELARALIEGGRFLWITDKLNREAKEIEGEEETEIEKIAGFPHLEEVGMIVSQCQI 338

Qy 347 EVLKHPSLGCFSVHCWNLSALESLAGVPVPAFPQWTDQMTNAKQVEDVWKSQVRYNE 406
Db 339 EVLHRAIGCGFLTHCGWSSSLESVLGVVPAFPWMSDQANAKLLEIEIKTGVRYRENS 398

Qy 407 DGVESEIEKRLIYVMDGEGEELRNKAKKWLAREAVKEGSSHKNLKAFI 461
Db 399 EGLVERGEIMKLEAVNEA--KSVLELRENAEKWLAREAVKEGSSDKKNVAFV 451

RESULT 9
Q23406 PRELIMINARY; PRT; 458 AA.
ID Q23406
AC Q23406;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)

```

```

DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE SIMILAR TO INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC eumollusca; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., STIEKEMA W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA PUGDOMENECH P., HATZIOPOULOS P., OBERMAIER B., DUESTERHOFF A.,
RA JONES J., PALME K., ANSGORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA SCHUELLER C., CHALATAIS N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97339; CAB10333.1; -.
DR MENDEL; 26727; Arath; 2543; 26727.
DR PROSITE; PS00375; UDEGT; 1.
DR PFAM; PF00201; UDEGT; 1.
KW Transferase.
SQ SEQUENCE 458 AA; 52005 MW; 67943A2F CRC32;

Query Match 42.4%; Score 1049.5; DB 10; Length 458;
Best Local Similarity 46.5%; Pred. No. 1.1e-68;
Matches 226; Conservative 78; Mismatches 115; Indels 67; Gaps 15;

Qy 4 PHVLTTPAQGHINPALQFAKNLVK--MGIEVTFSTSIYAQSRMDEKSIILNAPKGLNFI 61
Db 12 PHFLVTFPAQGHINPSLEAKRLAGTISGARVTFFAAISAYNR--MFSTENVETLIFA 70

Qy 62 PFSDFDEGFG-----DHSKDPVF--YMSQLKCGSETVKKIILTCEGNGPITCLLYSI 113
Db 71 TYSGDHDDGFKSAYSIDKSRQDATGNFMSEMRREGKTELLEEDNRKONRPFCTVYPTI 130

Qy 114 FLPAWAEVAREVHIPALLWSQATILDIYFNFHGYEKAMANESNDPNMSIQLPGLPL 173
Db 131 LLTWVLAEL-----LFSIFVHYFNGYDAISEMANTPSSSIKLPISLPL 174

Qy 174 ETRDPSFLPYGAGSLVALPPEKELIDTDAETTPKILVNTFDELEPEALNATIEGK 232
Db 175 TVRDIEFTSVSSNVAFL---LPAPREQIDSLKEEINPKILINTFOELEPEAMSSVPDNF 231

Qy 233 KFYGIGLIPSAFAGNDPDLDAFGDGLFQNSNDYMEWLNKSPNSVYIYFSGLSMNPIS 292
Db 232 KIVPVGPLL-----TLRTD--FSSKGEYIEWLDTKADSSVLYVSGTAVLSK 277

Qy 293 SQMEELSKGLIDIGRPFLWVI-----KENEKKEENKKGICL-----EELEKIGKIVP 341
Db 278 KQLVELCKALIOSRRPFLWITDKSYRNKDEQKEED-----CISFRELEDEIGNVYS 332

Qy 342 WCSQLEVLKHPSLGCFSVHCWNLSALESLAGVPVPAFPQWTDQMTNAKQVEDVWKSQV 401
Db 333 WCDQFVRLNHRSGFCVTHCGWNTLSLSVSGVPVPAFPQWDMNNAKLEDCWKTGVR 392

Qy 402 V--RINEDG--VVESEIEKRLIYVMDGEGEELRNKAKKWLAREAVKEGSSHKNL 457
Db 393 VMEKKEEGVYVVDSEIRRCIEEYNE--DKAEFRGNATRWKDLAAAVREGGGSFNHL 450

Qy 458 KAFDID 463
Db 451 KAFVDE 456

RESULT 10
Q23401 PRELIMINARY; PRT; 479 AA.
ID Q23401
AC Q23401;
DT 01-JAN-1998 (TReMBLrel. 05, Created)

```

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE INDOL-3-ACETATE BETA-GLUCOSYLTRANSFERASE
 GN F18019.7.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., STIEKENE W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,
 RA KREIS M., KAVANAGH T., ENIAN K.D., RIEGER M., JAMES R.,
 RA PUIGDOMENICH P., HATZPOULOS P., OBERMAIER B., DUESTERHOFF A.,
 RA JONES J., PALME K., ANGEORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
 RA SCHUELLER C., CHALWATZIS N.
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 297339; CAB10327.1; -;
 DR PROSITE; PS00375; UDPGT; 1.
 DR PFAM; PF00201; UDPGT; 1.
 KW Transferase.
 SQ SEQUENCE 479 AA; 53913 MW; 2ED16CCC CRC32;

Query Match 32.7%; Score 809.5; DB 10; Length 479;
 Best Local Similarity 38.4%; Pred. No. 3.4e-51;
 Matches 187; Conservative 83; Mismatches 158; Indels 59; Gaps 15;
 QY 5 HVILTFPAQGHINPAQFAKNLVKMGIEVTFSTSIYASRMDKSIILNAPKGL 58
 DB 8 HVMLVFPQGHVNFLLRLKLTASKGLAVTFVTEKPGKMKRQANKIDGVLYK-PVGL 66
 QY 59 NFIP---FSDFGDFHSDKDPVFNYSQLRKCGSETVKKIILTCSENGQPTICLLYSIFL 115
 DB 67 GFIREFFSDGFADDEKRFDFAPRPHLEAVGQBIKLVK--RYNKEPTVCLINAFV 124
 QY 116 PMAEVAAREVHIPALLMSOPATLIDYYFNHFGYKAMANESNDPNWSIOLPGLPLET 175
 DB 125 PWCVDAAELHIPSAVLWQSCACLTAYYYHRLVKFT--KTEPDLSVEIPCLPLKH 182
 QY 176 RDLPSFL--LPYGAKSRLVALPPKELIDTLDAETTPK---ILVTFDELE----- 222
 DB 183 DEIPSLPSSPYTAFGDI-----ILDLKRFENHKSPYLFIDTFRELEKDIMDHM 233
 QY 223 ----PEALNATEGYKPYGIGLIPSAFLGNDPLDASFGDLFQNSNDYMEWLNKPNSS 278
 DB 234 SOLCPQAI-----ISPGPLFKMA-----QTLSSDVKGDISEPACDMEWLDSEPPS 281
 QY 279 VYVTSFGSLNPSTSQMEETSKGLIDIGRPFLLWIKENKKEENKKGICIELEKIGK 338
 DB 282 VYVTSFGTIANLQGEELAHGVLSSGLVLMVVRPPMEGTFVEPHLP--RELEEKG 339
 QY 339 IVPWCSQLEVLKHPSLGCFYSHGWNLSALESACGVPVAFPPQWTDQMTNAKQVEDVWKS 398
 DB 340 IVEWCPQRLVAHPAICFLSHCGWNSTNEALTAGVPVVCVFPQWGDQVTDVYLADVFKT 399
 QY 399 GVRV--RINEDGVVESEIT--KRCIELVMDGEGEELRNKAKWELAREAVKGGSGSHK 455
 DB 400 GVRIGRGAEMIVSRVVAEKLEATV--GEKAVELRENARWRAREAAVADGGSSDM 457
 QY 456 NUKAFID 462
 DB 458 NFKFVD 464

RESULT 11
 Q22822
 ID O22822 PRELIMINARY; PRT: 449 AA.
 AC O22822;
 DT 01-JAN-1998 (Tremblrel. 05, Created)

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE INDOL-3-ACETATE BETA-GLUCOSYLTRANSFERASE ISOLOG.
 GN F18019.7.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-CV. COLUMBIA;
 RA ROUNSLY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
 RA SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
 RA VENTER J.C.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC002333; AAB64024.1; -;
 DR MENDEL; 26507; Arath; 3145; 26507.
 DR PROSITE; PS00375; UDPGT; 1.
 DR PFAM; PF00201; UDPGT; 1.
 KW Transferase.
 SQ SEQUENCE 449 AA; 50771 MW; 8B0DCE0A CRC32;

Query Match 31.4%; Score 777; DB 10; Length 449;
 Best Local Similarity 36.6%; Pred. No. 7.2e-49;
 Matches 173; Conservative 98; Mismatches 156; Indels 46; Gaps 16;
 QY 5 HVILTFPAQGHINPAQFAKNLVKMGIEVTFSTSIYASRMDKSIILNAPKGLNFI 64
 DB 7 HVLAVPYPTQGHITPFRQCKLRHFGKLTALTTFVNSINPD--LSGP--TSIATIS 62
 QY 65 DGFDEGDFHSDKDPV-FYMSQLKCGSETVKKIILTCSENGQPTICLLYSIFLPMAAEVAR 123
 DB 63 DGYDHGSETADSIDDYLDKFTSGSKTIADIIQKHTSDNPITCIVTDAFLPWALDVAR 122
 QY 124 EVHPSALLMSQATLIDYYFNHFGYKAMANESNDPNWSIOLP--GLPLETRDLPSF 181
 DB 123 EFLGVAFTFPQCAVNVVYLSYN-----NGSLQPLEELPFELEQLDLPSP 170
 QY 182 LLPYGAKSRLVALPPKELI--DTLDAETTPKILVNTFDELEPALNAIGYK----FY 235
 DB 171 ---FSVSGS----YPAYFEMVLOQFINFEKADFLVNSFOELE--LHENELWSKACPV 220
 QY 236 GIGLIPSAFLGNDPLDASFGDLFQNSND--YMEWLNKPNSSVYVTSFGSLMNPSSIS 293
 DB 221 TIGPTIPSIYLDQRIKSDTCGYDLNLFESKDDSPCINWLDTRPQGSVVVAVFGSMAQLTNV 280
 QY 294 QMEELSKGLIDIGRPFLLWIKENKKEENKKGICIELEKIGKIV-PWCSQLEVLKHP 352
 DB 281 QMEELASAVSNFS--FLWVYRSS-----BEKLPSTVETVKNKESLVKLSQLOVLSNK 334
 QY 353 SLGCFVSHGWNLSALESACGVPVAFPPQWTDQMTNAKQVEDVWKSQVVRVRI-NEDGVVE 411
 DB 335 AIGCFLTHCGWNSTNEALTGVPVNVAMPQWTDQPNNAKIYQDVWKAQVVKTERESGIK 394
 QY 412 SEETKRCIELVMDGEGEELRNKAKWELAREAVKGGSGSHKNAKAFIDV 464
 DB 395 REEIEFSIKEVHE--GERSKEMKNVKKWRDLAVKLSUNEGGSDTIDITFVSRV 446

RESULT 12
 O23402
 ID O23402 PRELIMINARY; PRT: 475 AA.
 AC O23402;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE INDOL-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

QC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., STIKENKA W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,
 RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
 RA PUIGDOMENECH P., HATZPOULOS P., OBERMAIER B., DUESTERHOFT A.,
 RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
 RA SCHUELLER C., CHALWATZIS N.,
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97339; CAB10328.1; -;
 DR PROSITE: PS00375; UDPGT; 1.
 DR PFAM: PF00201; UDPGT; 1.
 KW Transferase.
 SQ SEQUENCE 475 AA; 53495 MW; A87EAE5 CRC32;
 Query Match 30.9%; Score 766; DB 10; Length 475;
 Best Local Similarity 36.9%; Pred. No. 5e-48;
 Matches 175; Conservative 91; Mismatches 172; Indels 36; Gaps 14;
 QY 4 PHVILTFPAQCHINPALQFAKNLVKMGIEVTFSTS-----TYAQRMDKESILNAPKG 57
 DB 8 PHVILVSPQCGHISPLRLKGIILASKGLIVFTVEEPGKKHROANNIQDGLK-PVG 66
 QY 58 LNFIPFSGDFGDFHSDKDPVYFMSQLKRCGSETVKKIILTCSENGQIPITCLLYSIFLFW 117
 DB 67 LGFLRF-EFFEDGVYKEDFDLLQKSLVSKREIKNLVKYK--QPVRLINNAFVW 123
 QY 118 AAEVAREVHIPSAALLWSQATILDIYFNFHGYEKAMANESNDPNSIOLPGLP-LETR 176
 DB 124 VCDIAELQIPSAVLWVOSCAALAAIYYIHOLVKFT--ETEPEITVDVPPKPTLKH 181
 QY 177 DLPSFLLPYGAAGSL-RVALPPFKELIDTDAETTPKILVNTFDEPEALNAIE-----G 231
 DB 182 EIPSLFHPSSPLSSIGGTILQIKRLHKPF-----SVLIETFOELEKTDIDHMSQLCPQ 235
 QY 232 YKFGYGLPIPSAFGLGNDPLDASFGDLQNSNDYMEWLNSKPNSSVYISFGSLMNP 291
 DB 236 VFNPIGLPFTMA-----KTRSDIKGDISKPSDCLEWLDSEPSVYISFGTLFLK 290
 QY 292 ISQMBEISKGLIDIGRPFLWIKENKRGKEENKKGICIEBELEKIGKIYVPC-SQELVKH 351
 DB 291 QNOCIDEIAGILNSGLCLWLVRPPLEGALTEPHVLP--ELEKGIKIVECQOEKVLAH 348
 QY 352 PSIGCFVSHCGNSALESIACGVPVAPPOWTDQMTNAKOVEDYKSGVVR--INEDGV 409
 DB 349 PAVACFLSHCGWNSMTNALTSGVPVPCFPGQDQVTVNAVYMDVFTGLRSGRGSADRI 408
 QY 410 VSEET-KRCIELVMDGGEELKRNKAKKWLKELAREAVKGGSSHKLNKAFID 462
 DB 409 VPREEVAERLEATV--GEKAVELRENARWKAEASAVAYGCTSERNFQEFVD 460
 RESULT 13
 ID O48676 PRELIMINARY; PRT; 460 AA.
 AC O48676;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE F316.2 PROTEIN.
 GN F316.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;
 RA FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
 RA AU M., ARAJO R., BUEHLER E., DEMAR K., FENG J., KIM C., LI Y.,
 RA OJI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VYSOTSKAIA V.S.,
 RA YU G., ECKER J., THEOLOGIS A., DAVIS R.W.,
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002396; AAC00570.1;
 DR MENDEL: 27887; Arath:3145;27887.
 DR PROSITE: PS00375; UDPGT; 1.
 DR PFAM: PF00201; UDPGT; 1.
 SQ SEQUENCE 460 AA; 51002 MW; 02F08C19 CRC32;
 Query Match 30.8%; Score 762; DB 10; Length 460;
 Best Local Similarity 38.2%; Pred. No. 9.3e-48;
 Matches 187; Conservative 78; Mismatches 163; Indels 62; Gaps 17;
 QY 2 VQPHVILTFPAQCHINPALQFAKNLVKMGIEVTFSTSIIYAQRMDKESILNAPKGLAPI 61
 DB 8 VGHVVILPYPVQGHLPMPVQFAKRLVSKNVKVTIATTTASSITTPS-----LSVE 60
 QY 62 PFDGFD-----EGFDHSDKDPVYFMSQLKRCGSETVKKIILTCSENGQIPITCLLYSIFL 115
 DB 61 PISDGFDFPIGPGF--SVDT--YSEFKLNGSETLTLLIEFKSTDSPIDCLLYDSFL 116
 QY 116 PNAEAREVHIPSAALLWSQATILDIYFNFHGYEKAMANESN-----DPNWS-IOL 167
 DB 117 PMGLEVARSMELSAASFTNLTVCV-----LRKFSNGDFPLPADPNSAPFRI 165
 QY 168 PGLPLETRDLPSFL-----LPYGAGSLRVALPPFKELIDTDAETTPKILVNTFDELEP 223
 DB 166 RGLPSLSYDELPFSFVGRHMLTHPEHG--RVLLNQFP-----NHENADMLVNGFEGLE- 216
 QY 224 EALNAIEG-----YKFGYGLPIPSAFGLGNDPLDASFGDLFON-SNDYMEWLNSKPNSS 278
 DB 217 ETQDCENGESDAMKATLIGPMIPSAIYLDNRMDKDYGASLLKPISEKCEWLETKQAOS 276
 QY 279 VYVIFSGSLMNPISIQMBEISKGLIDIGRPFLWIKENKRGKEENKKGICIEBELEKIGK 338
 DB 277 VAFVSEGSFGILFEKQLAEVAIALQESDLNFWIKAEHAKLPE---GFVESTKDRAL 332
 QY 339 IYPCWSQLVLEKHPISGCFVSHCGNSALESIACGVPVAPPOWTDQMTNAKOVEDYKWS 398
 DB 333 LYSWCNQLVLEAHESIGCFVTHCGWNSVTEGLSLGVPMVGVQWSDQMDNAKEVEWVKV 392
 QY 399 GYVRVINEG--VVSSEETKRCIELVMDGGEELKRNKAKKWLKELAREAVKGGSSHK 456
 DB 393 GYRAK-EEAGEYIVKSEELVRLCKGVE-GESSVKIRESSKKWKLAVKAMSEGGSSDRS 450
 QY 457 LKAFIDVAK 466
 DB 451 INEFIESIGK 460
 RESULT 14
 ID O23400 PRELIMINARY; PRT; 484 AA.
 AC O23400;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DE INDOL-3-ACETATE BETA-GLUCOSYLTRANSFERASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., STIKENKA W., MURPHY G., WAMBUIT R., POHL T., TERRY N.,
 RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
 RA PUIGDOMENECH P., HATZPOULOS P., OBERMAIER B., DUESTERHOFT A.,
 RA JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,

RA SCHUELLER C., CHALWATZIS N.:
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97339; CAB10326.1; -
 DR PROSITE: PS00375; UDCPT; 1.
 DR PFAM: PF00201; UDCPT; 1.
 KW Transferase.
 SQ SEQUENCE 484 AA; 53876 MW; 88A0F3BB CRC32;

Query Match 30.7%; Score 761; DB 10; Length 484;
 Best Local Similarity 36.9%; Pred. No. 1.2e-47;
 Matches 178; Conservative 87; Mismatches 178; Indels 40; Gaps 15;

QY 5 HVLITTPAACHINPALQFAKNLVKMGIEVTF--STSIYAQSRMDKSKILNA---PKGLNF 60
 DB 13 HVLVYFQGGCHVNPPLRLGLKLSKGLLVTFVTELAKKMRQANKIVDGELRPVSGS 72
 QY 61 IFFSDGDEGF---DHSKDPVYMSOLRKCGETVKKIILTCSENGQPTICLLYSIFLP 116
 DB 73 IRE-EFDEEWAEDDDRRADFSLYIAHLESVGLREVSKLVRYREANEPVSCLINPFIP 131
 QY 117 WAAEVAREVHIPALLWSQPATILDIYFNFHGYEKAMAN--ESNDPNWSIOLPGLPILLE 174
 DB 132 WYCHVAEEFNPCAVLWQSCACFSAYH---YDGSVSFPFTEPELDVKLPVLPVK 187
 QY 175 TRDLPSPFLPYGAKGSLRVA-LPPFKELIDTLDAETTPKILVNTPEPEALNAIEGY- 232
 DB 188 NDEIFLHPFSRRTGFRQAILGQFNLSKSF-----CVLIDSPDSELEQVIDYMSSIC 241
 QY 233 KFYGIGPLIFSAFGLGNDPLDASFGDGLFQNSNDYKWLKNPSSVYVYISFGSLMPSI 292
 DB 242 PVKTVGPLFVKA-----RTVTSVDSGDICKSTDKCLEWLDSPKSSVYVYISFGVAYLKQ 296
 QY 293 SQMEISKGLIDIGRPPLVAVIKENEGKEENKKGCIIELEKI-----GRIVPWCQLE 347
 DB 297 EQIEEIAHGVLSKLSLFWIRPPHDLKVETHVLP--QELKESAKGKGMIVDMCQEQ 354
 QY 348 VLKHPSLGCFVSHCGWNSALESACGVPVVAFPQWTDQMTNNAKQVEDYVKSQVRY--RIN 405
 DB 355 VLSHPSVACVTHCGWNSTMSLSSGVPVVCQPMQDQVTDAYVILIDVFKTVGLRGAT 414
 QY 406 EDGVVSEDEI-KRCIELVMDGGKGBELRNKAKWKLAREAVKGGSSHKNLKAFFDDV 464
 DB 415 EERVVPREEVAEKLLEATV--GEKAEELRNKALWKAEAAVAPGGSSDKNREFVEKL 472
 QY 465 AKG 467
 DB 473 GAG 475

RESULT 15
 O22820 PRELIMINARY; PRT; 449 AA.
 AC O22820;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
 DE INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE ISOLG.
 GN F18019.5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsids.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
 RA SPRIGGS T.A., NASON T.M., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,

RA VENTER J.C.:
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002333; AAB64022.1; -
 DR MENDEL; 26506; Arath:3145;26506.
 DR PROSITE: PS00375; UDCPT; 1.
 DR PFAM: PF00201; UDCPT; 1.
 KW Transferase
 SQ SEQUENCE 449 AA; 50331 MW; FB3CC5C8 CRC32;

Query Match 29.6%; Score 733; DB 10; Length 449;
 Best Local Similarity 35.7%; Pred. No. 1.2e-45;
 Matches 168; Conservative 93; Mismatches 168; Indels 42; Gaps 16;

QY 5 HVLITTPAACHINPALQFAKNLVKMGIEV--TFSTSIYAQSRMDKSKILNAPKGLNPIP 62
 DB 7 HVLAVPFPFOCHITPQFCKRLHSGFKTHTLTFTIFNTIHLDPSS-----PISIAT 60
 QY 63 FSDGFDE-GFDHSDKDPVYMSOLRKCGETVKKIILTCSENGQPTICLLYSIFLPAAEV 121
 DB 61 ISDGYDQGGFSSAGSVPEYLQNEKTFGSKTVADIIRKHQSTDNPTICIVDSFPMALDL 120
 QY 122 AREVHIIPALLWSQPATILDIYFNFHGYEKAMANESNDPNWSIOLPGLPILLETDLPSF 181
 DB 121 AMDFGLAAAPFFTQSCA---VNYINLSY-----INNGSLTLPKIDLPLEQLDLPF 170
 QY 182 LLLPYGAK-GSLRVALPFPKELIDTLDAETTPKILVNTPEDEL---EPEALNAIEGYKFI 237
 DB 171 VTPGSHLAYFEMVLOQFTN-FDKADF-----VLVNSEPHDLHHEEELSKV--CPVLT 222
 QY 238 GPLIPSAFLGNDPLDASFGDGLFQNSNDY--EWLNSKPNSSVYVYISFGSLMPSISOM 295
 DB 223 GPTVPSMYLQOIKSNDYDLNLFDLKEALCTDMLDKRPEGSVYVYIAFGSMALSSQOM 282
 QY 296 BEISKGLIDIGRPPLVAVIKENEGKEENKKGCIIELEKIIV-PMCSOLEVLKHPSL 354
 DB 283 BEIASAISNFS--YLMWVRASESKLPP---GFLETVDKOKSLVLKWSPOLVLSNKAI 336
 QY 355 GCFVSHCGWNSALESACGVPVVAFPQWTDQMTNNAKQVEDYVKSQVRY--NEDGVVSE 413
 DB 337 GCFVTHCGWNSTMEGLSLGVPVVAFPQWTDQPMNAKIQDWKVGVRVKAESGICKRE 396
 QY 414 EIKRCIELVMDGGKGBELRNKAKWKLAREAVKGGSSHKNLKAFFDDV 464
 DB 397 EIEFSIKVME-GEKSKEMKENAGKWRDLAVKLSLSEGGSTDININEPVSKI 446

Search completed: August 1, 2000, 21:31:44
 Job time: 6073 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 1, 2000, 19:51:36 ; Search time 59.57 Seconds
(without alignments)
120.431 Million cell updates/sec

Title: 05-09-147-955-12
Perfect score: 2475
Sequence: 1 MVQPHVLTTPAQGHINPA.....EGSSHKNAKAFIDVAKGF 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/PCRU.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	750.5	30.3	471	2	US-08-466-583-2
2	750.5	30.3	471	4	PCT-US95-07820-2
3	481	19.4	471	3	US-09-106-464-2
4	443.5	17.9	488	2	US-08-797-226-2
5	359	14.5	131	2	US-08-466-583-5
6	339	14.5	131	4	PCT-US95-07820-5
7	286	10.7	131	2	US-08-466-583-6
8	266	10.7	131	4	PCT-US95-07820-6
9	225	9.1	63	2	US-08-466-583-8
10	225	9.1	63	4	PCT-US95-07820-8
11	189	7.6	531	4	PCT-US92-00282-6
12	183	7.4	534	4	PCT-US92-00282-4
13	173	7.0	529	4	PCT-US92-00282-7
14	173	7.0	533	4	PCT-US92-00282-3
15	168	6.8	531	4	PCT-US92-00282-5
16	128.5	5.2	506	5	5180581-2
17	123	5.0	56	2	US-08-466-583-4
18	123	5.0	56	4	PCT-US95-07820-4
19	114	4.6	74	4	PCT-US92-00282-24
20	112	4.5	409	2	US-08-924-254-2
21	112	4.5	409	3	US-09-120-249-2
22	107.5	4.3	414	2	US-08-750-524-1
23	106	4.3	3443	2	US-08-416-603-2
24	104.5	4.2	510	1	US-08-278-635B-4
25	102.5	4.1	511	3	US-08-464-258B-4
26	100	4.0	58	2	US-08-466-583-9
27	100	4.0	58	4	PCT-US95-07820-9
28	98	4.0	52	2	US-08-466-583-7

29	98	4.0	52	4	PCT-US95-07820-7	Sequence 7, Appli
30	97.5	3.9	529	1	US-08-496-855A-2	Sequence 2, Appli
31	96	3.9	528	2	US-08-466-589-2	Sequence 2, Appli
32	96	3.9	528	2	US-08-700-636-2	Sequence 2, Appli
33	96	3.9	528	3	US-08-467-574-2	Sequence 2, Appli
34	94	3.8	780	2	US-09-018-760-4	Sequence 4, Appli
35	92.5	3.7	2391	2	US-08-446-855A-2	Sequence 2, Appli
36	89.5	3.6	1031	1	US-07-689-008-2	Sequence 2, Appli
37	88	3.6	1220	2	US-08-843-530B-36	Sequence 36, Appli
38	85	3.4	775	2	US-08-966-388-4	Sequence 4, Appli
39	85	3.4	775	3	US-09-188-403-4	Sequence 4, Appli
40	84	3.4	1022	1	US-08-271-364A-8	Sequence 8, Appli
41	84	3.4	1022	2	US-08-222-715B-27	Sequence 27, Appli
42	83.5	3.4	408	2	US-08-926-358-2	Sequence 2, Appli
43	83.5	3.4	408	2	US-09-120-053-2	Sequence 2, Appli
44	83.5	3.4	552	3	US-09-111-752-5	Sequence 5, Appli
45	83.5	3.4	897	1	US-07-960-389-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-466-583-2
; Sequence 2, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeizzen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/365,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-583-2

Query Match 30.3%; Score 750.5; DB 2; Length 471;
Best Local Similarity 35.5%; Pred. No. 1.2e+68;
Matches 172; Conservative 85; Mismatches 188; Indels 39; Gaps 13;

```
QY 4 PHVILTTTPOAGHINPALQFAKNLVKMGIEVTESTSIYAQSRMD---EKSILNAPKGLNF 60
Db 3 PHVLVPPFGGHNPNVQFAKRLASKGVATTLVTRFIQRTADVDHAPAMVEA-----56
QY 61 IPFSDGDE-GFDHSDKDPVFYMSQLKRCGSETVKKIILTCSENGQPIITCLLYSIFLPMAA 119
Db 57 --ISDGHDEGGFASAGVAEYLEKQAAAAASLASLVEARASSADAFCTCVVYDSYEDWVL 114
QY 120 EVAREVHIPSSALLNSQATILDIYFNFHCY-----EKAMANESNDPNWSIQLP---GLP 171
Db 115 PVARRMGILPVPFSTQSCAVSYVYHFSGRLAVPPGAADSGDGAGAAALSEAFILGLP 174
QY 172 LLETDLPSFLPYGAKGSLRVLPPELIDTLDATTPKILVNTFDELEPEALNAIEG 231
Db 175 EMERSELPFVDRHGPT--IAMQAIKQF---AHAGKDDWVLFNSFEETEVLGLTK 229
QY 232 Y-KFYGIGPLIP---SAFLGGNDPLDASGGDLFONSDYMEWLNKSPNSVVIISFGS 286
Db 230 YLKARAIGPCVPLPTAGTAGANGRI--TYCANLVKPEDACTKWLDTKPDRSVAVVSFGS 287
QY 287 LMNPSISOMEBISKGLIDIGRPFLWIKENEKKEENKKGICIELEKIG--KIVPMCS 344
Db 288 LASLGNQAKKEELARGLLAAGKPELWVY---RASDEHQVPVRYLLAEATATGAAMVVPWCP 343
QY 345 QLEVLKHPSLGCFYSHCGWNSALESACGVPVYAFPOWTDQMTNAKQVEDVHKSGVRVRI 404
Db 344 QLDVLHAPVAGCFVTHCGWNSTLEALSFGVPVYAMALWTDQPTNARNVELANGAGVARR 403
QY 405 NED-GVVESEBIKRCIELVMDGGEKEELKNAKKWELAREAVKEGSSSHKNAKAFIDD 463
Db 404 DAGAGVFLRGEVERCVRAVMDGGEAASAAKAGWRDRARAAPVGGSDRNLDDEFVQF 463
QY 464 YAKG 467
Db 464 VRAG 467

RESULT 2
PCT-US95-07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
```

```
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07820-2

Query Match 30.3%; Score 750.5; DB 4; Length 471;
Best Local Similarity 35.5%; Pred. No. 1.2e-68;
Matches 172; Conservative 85; Mismatches 188; Indels 39; Gaps 13;

QY 4 PHVILTTTPOAGHINPALQFAKNLVKMGIEVTESTSIYAQSRMD---EKSILNAPKGLNF 60
Db 3 PHVLVPPFGGHNPNVQFAKRLASKGVATTLVTRFIQRTADVDHAPAMVEA-----56
QY 61 IPFSDGDE-GFDHSDKDPVFYMSQLKRCGSETVKKIILTCSENGQPIITCLLYSIFLPMAA 119
Db 57 --ISDGHDEGGFASAGVAEYLEKQAAAAASLASLVEARASSADAFCTCVVYDSYEDWVL 114
QY 120 EVAREVHIPSSALLNSQATILDIYFNFHCY---EKAMANESNDPNWSIQLP---GLP 171
Db 115 PVARRMGILPVPFSTQSCAVSYVYHFSGRLAVPPGAADSGDGAGAAALSEAFILGLP 174
QY 172 LLETDLPSFLPYGAKGSLRVLPPELIDTLDATTPKILVNTFDELEPEALNAIEG 231
Db 175 EMERSELPFVDRHGPT--IAMQAIKQF---AHAGKDDWVLFNSFEETEVLGLTK 229
QY 232 Y-KFYGIGPLIP---SAFLGGNDPLDASGGDLFONSDYMEWLNKSPNSVVIISFGS 286
Db 230 YLKARAIGPCVPLPTAGTAGANGRI--TYCANLVKPEDACTKWLDTKPDRSVAVVSFGS 287
QY 287 LMNPSISOMEBISKGLIDIGRPFLWIKENEKKEENKKGICIELEKIG--KIVPMCS 344
Db 288 LASLGNQAKKEELARGLLAAGKPELWVY---RASDEHQVPVRYLLAEATATGAAMVVPWCP 343
QY 345 QLEVLKHPSLGCFYSHCGWNSALESACGVPVYAFPOWTDQMTNAKQVEDVHKSGVRVRI 404
Db 344 QLDVLHAPVAGCFVTHCGWNSTLEALSFGVPVYAMALWTDQPTNARNVELANGAGVARR 403
QY 405 NED-GVVESEBIKRCIELVMDGGEKEELKNAKKWELAREAVKEGSSSHKNAKAFIDD 463
Db 404 DAGAGVFLRGEVERCVRAVMDGGEAASAAKAGWRDRARAAPVGGSDRNLDDEFVQF 463
QY 464 YAKG 467
Db 464 VRAG 467

RESULT 3
US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 6011145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Ghangas, Gurdev S.
; APPLICANT: Kua, Jian-Ping
; APPLICANT: Fannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc:Fatty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106.464
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US 60/055,554
APPLICATION NUMBER: US 60/055,554
FILING DATE: 13-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Spector, Eric S.
REGISTRATION NUMBER: 22495
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-415-1500
TELEFAX: 703-415-1508
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-464-2

Query Match 19.4%; Score 481; DB 3; Length 471;

Best Local Similarity 28.9%; Pred. No. 6.3e-41; Mismatches 167; Indels 128; Gaps 22;

Matches 153; Conservative 81; Mismatches 167; Indels 128; Gaps 22;

QY 1 MVQPHVLTTPAQGHINPALQFAKNLVKMGIEVTFST-----SIYAQSRMDEK-----49

DB 2 MGQLHFFPPMAQGHIMPTLDMAKLVACGVKATITTPINESVFSKATERNKHLGIEI 61

QY 50 --SILNAP-----KGLNIFPDSFGDFGDKDPVYMSQLKRCGSTVTKKIL 97

DB 62 DIRLLKPAKENDLPEDCERLDLPDSD-----KLPNFKAAAM--KDFEELIG 110

QY 98 TCSGNGQPTCLYSIFLPMAAEVAREVHIPSALLWSQPATILDIYFNFHG-----149

DB 111 ECRPD-----CLVSDMFLPMTTDSAAKFSIPRIV-----FHGTSYFALCV 150

QY 150 -----YEKAWANESNDPNWSIOLPGLPLETRDLPSELPLPYGAKGSLRVLPFKELIDT 204

DB 151 GDSIRKPKFNKVNSSD-----TETFPVPDF--PHEIR-LTRTQLSFEQ-----191

QY 205 LDAET--TPKI-----LVNFDLEP--EALNATEGYKFGYIGIGLIPSAFL 246

DB 192 SDETCMAPKRAVRESDAKSYGVFNFSYELESYVEHYTKVGRKNWAIGPL---SLC 248

QY 247 GGNDPLDASFGDLFQNSNDYMEWLNKPNSSVYVYISFGSLMNPISOMEEISKGLIDIG 306

DB 249 NRDIETKAERKSSIDEHACULWLDSSKSSIVYVCGSTADFTTAQMQLAMGLEASG 308

QY 307 RPELWIKE-NE-----KGREENKKLGIEELEKIGKIVPWCSOLEVLKHPISLGCFSVSHC 361

DB 309 QDFIWRITGNEDWLPGEFEERTKEGLI-----LRGAPQVLLDHEAIGAFVTHC 360

QY 362 GWSALESACGVVPAFQWTDQMTNAKQVEDVWKGSRV-----RINEDGVVSEBEI 415

DB 361 GWNSTLEGISAGVPMLTWVFAEQFNEKLVTEYMRGAGVSGKQWKRTASEG-VKREAI 419

QY 416 KRCIELVMDGGKGEELRKNNAKWEKELAREAVKKGSGSHHKNLAKAFIDDV 464

DB 420 AKAIKRVMA-ASETEGFRSRAKEYEMAREALEEGGSSYNGWATLIQDI 467

RESULT 4

US-08-797-226-2

; Sequence 2, Application US/08797226

; Patent No. 5959180

; GENERAL INFORMATION:

; APPLICANT: MOEHS, CHARLES P

; APPLICANT: ALLEN, PAUL V

; APPLICANT: ROCKHOLD, DAVID R

APPLICANT: STAPLETON, ANDREW
APPLICANT: GARBARINO, JOAN E
APPLICANT: FRIEDMAN, MENDEL
APPLICANT: BELKNAP, WILLIAM R
TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
TITLE OF INVENTION: GLYCOALKALOIDS IN SOLANACEOUS PLANTS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: NANCY J. PARSONS
STREET: 800 BUCHANAN ST.
CITY: ALBANY
STATE: CA
COUNTRY: USA
ZIP: 94710
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/797.226
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PARSONS, NANCY J
REGISTRATION NUMBER: 40,364
REFERENCE/DOCKET NUMBER: 0011.97
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 559-5731
TELEFAX: (510) 559-5777
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 488 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-797-226-2

Query Match 17.9%; Score 443.5; DB 2; Length 488;

Best Local Similarity 27.7%; Pred. No. 4.8e-37;

Matches 140; Conservative 95; Mismatches 191; Indels 79; Gaps 21;

QY 5 HVLTTPAQGHINPALQFAKNLVKMGIEVTFST-----IYAQSRMDEKSIILNPKGLN 59

DB 12 HVLFLPLSAGHIPILVNAARLFPASRGVKATITTPHNALLFRSTIDDDVRISGFPISIV 71

QY 60 PIPFSD---GFDEG---FDHSDP-----VFY-MSQLKRCGSTVTKKILTCSENGOPIT 107

DB 72 TIKPPSAEVLPGIEFSNATSPEMPHKIFVALSLQLKPMEDKIREL-----RP-D 122

QY 108 CLLYSIFLPMAAEVAREVHIPSALLWSQPATILDIYFNFHGYEKAMANESNDPNWSIOL 167

DB 123 CIFSDMVFPVTDIADLHIP-RILYNLSAYMCYSIMINLVY-RPHQPNLDSQSFFV 180

QY 168 PGLPLETRDLPSELPLPYGAKGSLRVA---LPPFKELIDTLD--AETPKILVNTDPELE 222

DB 181 PGLP-----DEIKFKLSQ-LTDDLKSSDDOKTVFDELLEQVEDSEERSYGVHDTFYELE 234

QY 223 P---EALNATEGYKFGYIGIGLIPSAFLGNDPDLFASFGDLFQNSND-YMEWLNKSPNS 278

DB 235 PAYDYVYQKLKPKCWHFGPLSHFASKIRSKELISE-----HNNELVIDMLNAOKPKS 288

QY 279 VVYISFGSLMNPISOMEEISKGLIDIGRPFLLWIKENEK-----GKEENKKGCCIE 331

DB 289 VLXVSGSMARFPESQLNEIAQALDASNVYFIFVLRPNEETASWLPVGNLEDKTKGLY- 347

QY 332 ELEKIGKIVPWCSOLEVLKHPISLGCFSVSHCGWNSALESIACGVVPAFQWTDQMTNAKQ 391

DB 348 -----IKGWVPQLTMEHSATGGFMTHCGTNSVLEAITFGVPMITWPLVADQFYNEKV 400

QY 392 VE-----DVKSGVRVRINEDGVVSEIEIKRCIE--LVMDGGKGEELKKNNAKV 439

```
Db 401 VELRGKIGKIDVWNGEIEI---TGPVIESAKIREARLWISNGSEIINIRDRVWAM 457
QY 440 KEAREAVKEGSSHKHAKAFIDV 464
Db 458 SKMAQATNEGSSWNNLTALIOHI 482

RESULT 5
US-08-466-583-5
; Sequence 5, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; US-08-466-583-5

Query Match 14.5%; Score 359; DB 2; Length 131;
Best Local Similarity 48.9%; Pred. No. 2.7e-29;
Matches 65; Conservative 26; Mismatches 36; Indels 6; Gaps 2;

QY 269 EWLNSKPNSSVYVIFSGSLMNPISQMEIEISKGLIDIGRPFLLWIKENEKKEENKKLG 328
Db 3 KWLDTKPRSVAYVYVSGSLASIGNAQKEELARGLLAAGKPFLLWV----RASDEHQVPRY 58
QY 329 CIEELEKIG--KIVPMCSQLEVLKHPISLGCFSVHCGWNSALESACGVPVYVAFPOWTDQM 386
Db 59 LLAETATGAAMVYVPMCPQLDVLPAHVAFCVTHCGWNSTLEALSFGVPMVAMALWTDQP 118
QY 387 TNAKOVEDVWKSQ 399
Db 119 TNARNVELAWGAG 131

RESULT 7
US-08-466-583-6
; Sequence 6, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
```

```
RESULT 6
PCT-US95-07820-5
; Sequence 5, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szerszen, Jędrzej B.
; APPLICANT: Szczygłowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; PCT-US95-07820-5

Query Match 14.5%; Score 359; DB 4; Length 131;
Best Local Similarity 48.9%; Pred. No. 2.7e-29;
Matches 65; Conservative 26; Mismatches 36; Indels 6; Gaps 2;

QY 269 EWLNSKPNSSVYVIFSGSLMNPISQMEIEISKGLIDIGRPFLLWIKENEKKEENKKLG 328
Db 3 KWLDTKPRSVAYVYVSGSLASIGNAQKEELARGLLAAGKPFLLWV----RASDEHQVPRY 58
QY 329 CIEELEKIG--KIVPMCSQLEVLKHPISLGCFSVHCGWNSALESACGVPVYVAFPOWTDQM 386
Db 59 LLAETATGAAMVYVPMCPQLDVLPAHVAFCVTHCGWNSTLEALSFGVPMVAMALWTDQP 118
QY 387 TNAKOVEDVWKSQ 399
Db 119 TNARNVELAWGAG 131

RESULT 7
US-08-466-583-6
; Sequence 6, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
```

```

ADDRESSSEE: Greenlee & Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07820
FILING DATE: 19-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,427
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 11-94B PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 131 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
PCT-US95-07820-6

Query Match 10.7%; Score 266; DB 4; Length 131;
Best Local Similarity 38.7%; Pred. No. 9,8e-20;
Matches 53; Conservative 22; Mismatches 50; Indels 1

QY 268 MEWLNSKPSNVVTIFSGSLMNPSISQMEIEISKGLIDIGRPLVWIKENE-----
      : || : | | | | | | | : : : : | | | | | : | :
Db 2 LAWLGRQPARGAVVSFGTACPRPELRELAAGLEDSCAPFLMSLRDSSWPHLP
      : || : | | | | | | | : : : : | | | | | : | :
QY 323 ENKKLGIEELEKIGTKIVPCSOLEVLIKHPISGCFVSHCGWNLSALESLACGPVV
      : || : | | | | | | | : : : : | | | | | : | :
Db 62 RAAGTGS-----GLVVPNAPQAVILRHSPSVGAFTVTHAGNASVLEGLSGGPVMA
      : || : | | | | | | | : : : : | | | | | : | :
QY 383 TDQWTNAKQVEDVKXSG 399
      || || : | | | |
Db 115 GDQRNARSVAHVWGFG 131
      || || : | | | |

RESULT 9
US-08-466-583-8
Sequence 8, Application US/08466583
Patent No. 591998
GENERAL INFORMATION:
APPLICANT: Bandurski, Robert S.
APPLICANT: Szeszen, Jędrzej B.
APPLICANT: Szczygłowski, Krzysztof
TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
TITLE OF INVENTION: and Plant Growth.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee & Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303

```

BEST LOCAL SIMILARITY 23.9%; Pred. NO. 8.3e-11;
 Matches 109; Conservative 53; Mismatches 188; Indels 96; Gaps 22;

NAME: FELDER, DONALD M.
REGISTRATION NUMBER: 33,878

QY 13 AOGHINPALQAKNL-VKMGIEVTFSTSIYA-----QSRMDEKSILNAPKGLNIPFSDGF 67
Db 50 ARGH--QAVVLAPEVTHMGSGDFDTLQYAPPYTKBEYOREILGNNAKKGFEFOHFVKTF 107
QY 68 DEGFDSKDPVYMSQLRCKGSETVKKIILCSENGOPITCLLYSIFLPWAAEVAEVEVHI 127
Db 108 FETMASIKR--FFDLVANSCAALLHNKTLIQ-QLNSSSFDVLTDPVPCGALLAKYLQI 164
QY 128 PSA-LLWSOPARTILDIYFNFGYKAMANESNDPNKSIQLPGLPLET-----RDLPS 180
Db 165 PAVFELRSVPCGI-----DYATQC-----PKPSSYIPNLLTMLSDHMTFLORVKN 210
QY 181 FLPLGAKSLRVALPPFKELI-DPLDAETPKILVNTFDELEPEALNAIEGKFKYG--- 236
Db 211 MLYPLTLKYLCHLSITPYESLASSELLQREMS---LV-----EVLSHASVWLFGRGDFV 259
QY 237 ---IGPLPS-AFLG-----NDPLDASFGDLQNSNDYMEWLNKSPSSVYISFGSL 287
Db 260 FDPRPIMNWFPGINGCIVKPL-----SOEFAYVNASGEHGIYVFSLSGM 308
QY 288 MN--PSISQMEELSKGLIDIGRPFLLWIKENEKKEENKGLGCIIELEKIGKIVPWCQ 345
Db 309 VSEIPEKKAM-ETAEALGRIPOTLLW----RYTGTTPSN-----LAKNTILVKNLPQ 355
QY 346 LEVLKHPSLGCFVSHCGWNSALESACGVPVVAPOWTDQMTNAKOVEDVWKSVRVIRN 405
Db 356 NDLLGHPKARAFITHSGSGHIGEGICNGVPVWMPFLQDQMDNAKRME---TRGAGVTLN 412
QY 406 EDGVESEETKRCIELVMDGKGKBEELRNKAKWKE 441
Db 413 -----VLEMTADLENALKTVINNKSYKE 436

RESULT 12

PCT-US92-00282-4

; Sequence 4, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 534 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

PCT-US92-00282-4

Query Match 7.4%; Score 183; DB 4; Length 534;
Best Local Similarity 22.0%; Pred. No. 3.5e-10;
Matches 105; Conservative 71; Mismatches 163; Indels 138; Gaps 23;

QY 13 AOGH-----INPALQAKNLVKMGI--EVTFTSIYA-----QSRMDEKSILNAPKGLNIP 62
Db 53 ARGHQAVALTPE-----VNMHIKEKEFTLTAYAVPWTQKEFRVTL----- 94
QY 63 FSDGFDEGDSKDPVYMSQ-----LRKCGSETVKKIILCSENGOPITCLLY 111
Db 95 ---GYTQGFETEHLKRYSRMAIMNVSALHRCVVELLHNEALIRHLNATSFVLT 151
QY 112 SIFLPWAAEVAEVEVHIPLSALLSQPATILDIYFNFGYKAMANESNDPNKSIQLPGL- 170
Db 152 DPVNLGCAVLAKTSLIPAVFFWRYPCDLD-----FKG-----TQCPNPSYIPKLL 198
QY 171 -----ELLETRDLPSPFLPYGAKSLRVALPPFKELIDTLDATTPK 212
Db 199 TTNSDHMTFLORVKNMLYPLALSICYHTFSAPYAS-----LASELFQREVSVDLVSYAS 253
QY 213 ILVNTFDELEPEALNAIEGKFKYGIGPLPS-AFLGGNDPLDASFGDLQNSNDYMEWL 271
Db 254 VWLFGRD-----FVMDYP-----RPIMPNVFIGG---INCANGKPL---SOEFAYI 295
QY 272 NSKPNSSVYISFGSLMNPISQMEE-----ISKGLIDIGRPFLLWIKENEKKEENKK 326
Db 296 NASGEHGIYVFSLSGM-----VSEIPEKKAMATADALGKIPOTVLMW----RYTGTTPSNLA 347
QY 327 LGCTEELEKIGKIVPWCQSOLEVLKHPSLGCFVSHCGWNSALESACGVPVVAPOWTDQ 386
Db 348 NWI-----LVKWLQNDLLGHPWTRAFITHAGSHGVYESICNGVPVWMPFLQDQ 399
QY 387 TNAKQVEDVWKSVRVIRNEDGVVE--SEETKRCIELVMDGKGKBEELRNKAKWKE 441
Db 400 DNAKRME---TKGAGVTLN---VLEMTSEDLNALKAVI-----NDKSYKE 439

RESULT 13

PCT-US92-00282-7

; Sequence 7, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH

;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 529 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US92-00282-7

Query Match 7.0%; Score 173; DB 4; Length 529;
Best Local Similarity 22.2%; Pred. No. 3.7e-09;
Matches 98; Conservative 57; Mismatches 149; Indels 138; Gaps 19;

QY 91 TVKIIITCSENGQITCLLYSIFL-----PWAAEVAAREVH----- 126
DB 40 SKKEIVELSERGHDIWVPEYNLLGSKYRKSPVPYNLEELTRYRSFGNHEFA 99
QY 127 IPSALL-----WSOPATILDIYFNPHGYEKAMANES-----NDPNWSIQLP-G 169
DB 100 ASSFLMAPLREYRNMIVIDMCFQCSLLKDSATLSFLRENQFDALFTDP-----AMPCG 155
QY 170 LPLETRDLPSFLPYGAKSLR-----VALPPPKELIDTDAETTKILVNTFDE 220
DB 156 VILAELKLSIYLFRCFPCSLHIGQSPSPVSPRYFKFS--DHMTFFQRLANFI-- 211
QY 221 LEPEALNAIEGYKPYGI-----GRLIP 242
DB 212 -----ANILENYLHCLYSKYELASDLLKRDVSLPALHONSLWLLRYDFVFYPRPVPMP 265
QY 243 S-AFLGNDPLDASFGDLFONSDYMEWLNKPNSSVYISFGSLMN--PSISOMEEIS 299
DB 267 NWIFIGT---NCKKGNL---SQEFAVYNASGEHGVVFSLSKWSSEIPEKKAM-EIA 319
QY 300 KGLIDIGRPFLWIKENEKGBENKKGICIELEKIGKIVPWCQLEVLKHPSLGCFVS 359
DB 320 EALGRIPQTLW----RYTGRPSN-----LAKNTILVKWLPONDLLGHKPARAFIT 367
QY 360 HGWNSALESACGVPVAPFPQMTQDNNAKQVEDVWKSQVVRINEDGVVESEIEKRCI 419
DB 368 HSGSHGIEGICNGVPMVMPPLFGDQMDNAKRME---TKGAGVTLN-----VL 412
QY 420 ELYMDGGEKGEELKRNKKWKE 441
DB 413 EMTADDLENALKTVINNKSYKE 434

RESULT 14
PCT-US92-00282-3
;; Sequence 3, Application PC/TUS9200282
;; GENERAL INFORMATION:
;; APPLICANT: OWENS, IDA S.
;; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
;; TITLE OF INVENTION: THEREIN.
;; NUMBER OF SEQUENCES: 40
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSEE: CUSHMAN DARB & CUSHMAN
;; STREET: 1615 L STREET, N.W.
;; CITY: WASHINGTON
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20036-5601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US92/00282
;; FILING DATE: 19920110
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:

;; NAME: SCOTT, WATSON T.
;; REGISTRATION NUMBER: 26581
;; REFERENCE/DOCKET NUMBER: 91532-PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-861-3000
;; TELEFAX: 202-822-0944
;; TELEX: 6714627 CUSH
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 533 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US92-00282-3

Query Match 7.0%; Score 173; DB 4; Length 533;
Best Local Similarity 23.1%; Pred. No. 3.7e-09;
Matches 87; Conservative 64; Mismatches 126; Indels 100; Gaps 19;

QY 93 KKIITCSENGQITCLLYSIFLPWAAEVAAREVHIPSALLWSQPATILDIYFNPHGYEK 152
DB 134 KELMASLAESFDV--MLTDPFLPCSPIVAQVLSLPTV-----PFLHALPC 177
QY 153 AMANESND-PNWSIQLPGLPLETRDLPSF-----LLPYGAKSLRVALPPKELI-D 203
DB 178 SLEFEATQCPNPFYSVP-RPLSSHDMHTFLQRVKNMLTAFSQNFLCDVYVSPYATLASE 236
QY 204 TLDATTPKILVNT-----FDELEPEALNATEGYKFGYIGPLIPS-AFLGNDPLD 253
DB 237 FLQREVTVQDLLSSASVWLFERSDFVKDYP-----PIMPMVYFVGIN--- 279
QY 254 ASFGGDLFQN--SDNYMEWLNKPNSSVYISFGSLMNPISOMEE-----ISKGLIDIG 306
DB 280 -----CLHONPLSQEFAYNASGEHGVVFSLSGM---VSEIPEKKAMAIADALAKNP 330
QY 307 RPELWIKENEKGBENKKGICIELEKIGKIVPWCQLEVLKHPSLGCFVSHCGWNSA 366
DB 331 QTVLW----RYTGRPSNLANTTI-----LVKWLPPONDLLGHPTRAITHAGSHGV 378
QY 367 LESLACGVPVVPVAPFPQMTQDNNAKQVEDVWKSQVVRINEDGVVE--SEEKRCIELVMD 424
DB 379 YESICNGVPMVMPPLFGDQMDNAKRME---TKGAGVTLN---VLEMTSEDLENALKAVI- 431
QY 425 GGEKGEELKRNKKWKE 441
DB 432 -----NDKSYKE 438

RESULT 15
PCT-US92-00282-5
;; Sequence 5, Application PC/TUS9200282
;; GENERAL INFORMATION:
;; APPLICANT: OWENS, IDA S.
;; APPLICANT: RITTER, JOSEPH K.
;; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
;; TITLE OF INVENTION: THEREIN.
;; NUMBER OF SEQUENCES: 40
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSEE: CUSHMAN DARB & CUSHMAN
;; STREET: 1615 L STREET, N.W.
;; CITY: WASHINGTON
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20036-5601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US92/00282

```
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-00282-5

Query Match          6.8%; Score 168; DB 4; Length 531;
Best Local Similarity 20.0%; Pred. No. 1.2e-08;
Matches 92; Conservative 67; Mismatches 134; Indels 166; Gaps 20;

QY 19 PALQPAKLVKMGIEVTFSTIYAQSRNDEKSIINAPK----- 56
Db 108 POTEYRNNMIVIGLYF-----INCSLQDRDITLNFYKSKFDALFTDPALPGCVILAEY 162
QY 57 -GLNFIFFSDGF-----DEGDFHSKDPVFMNSQLRKCSETVVKRIILTCSENGQPIITCLLY 111
Db 163 LGLPSVYLFGRFPSCLEHTFSKSPDPVSY---IPRCYTKFSDHMTFS---QRVANFLV 214
QY 112 SIFLPAEVAEAREVHIPSAIILWSQPATILDIYFNFGYCEKAMANESNDPNKSIQLPGLP 171
Db 215 NLEPI-----LFYCLFSKYEK----- 231
QY 172 LLETDLPSFLPYGAKGLRVALPPFKELIDTLD-AETTPKILVNTFDELEPEALNAIE 230
Db 232 -----LASAVLKRDVDITLSEVSWLLRYDFVLEVP----- 264
QY 231 GYKFGIGPLIPS-AFLGGNDPLDASFGGDLFONSNDYMEMLNSKPNSSVWYISFGSLMN 289
Db 265 -----PVNPNWFIIG---INCKKKDL---SQEPEAYINASGEHGIIVFSLGSM-- 308
QY 290 PSISQHEE-----ISKGLIDIGRPFLWVIKENEKKEENKILGCIIELEKIGKIYPMCS 344
Db 309 --VSEIPEKKAMAIADALGKNPQTVLW----RYTGTRPSNLANNI-----LVKWL 354
QY 345 QLEVLKHPISLGCFSVSGWNSALESAGVPVAFPOWTDQMTNAKOVEDYWKSGYVRRI 404
Db 355 QNDLLGHPWTRAFTHAGSHGVYESICMGVPVWNPPLFGQMDNNAKRM---TKGAGVTL 411
QY 405 NEDGVVE--SEEIKRCELVDGGEELRKNKAKWKE 441
Db 412 N---VLEMTSEDLNALKRAVI-----NDKSYKE 436
```

Search completed: August 1, 2000, 19:51:38
Job time: 18760 sec

THIS PAGE BLANK (uspto)

Result No.	Score	Query Match	Length	DB ID	Description
1	101.8	6.1	1731	4	US-08-466-583-1
2	101.8	6.1	1731	6	PCR-US959-07920-1
3	62.6	3.7	7218	1	US-08-232-463-14
4	48.6	2.9	1669	3	US-08-522-421-1
5	45.4	2.7	1607	4	US-08-797-226-1
6	45.4	2.7	1627	5	US-09-106-464-1
7	40	2.4	405	4	US-08-975-316-32
8	38.8	2.3	4615	1	US-08-188-582-1
9	38.8	2.3	4615	1	US-08-646-715-1
10	38.4	2.3	1718	3	US-08-379-482A-2
11	37	2.2	1858	4	US-08-309-985C-11
12	37	2.2	53526	5	US-08-658-136-2
13	37	2.2	53577	5	US-08-658-136-1
14	36	2.2	251	3	US-08-623-906A-16
15	36	2.2	260	3	US-08-520-678A-29
16	36	2.2	2621	4	US-08-553-619B-8
17	35.6	2.1	806	5	US-08-906-769-120
18	35.6	2.1	1582	5	US-08-543-196B-10
19	35.6	2.1	1582	5	US-08-545-196B-12
20	35.4	2.1	855	7	5185441-40
21	35.4	2.1	855	7	5253394-3
22	35.4	2.1	863	1	US-07-940-861-11
23	35.4	2.1	863	2	US-08-459-512-11
24	35.4	2.1	863	4	US-08-459-657-11
25	35.4	2.1	863	4	US-08-460-132-11
26	35.4	2.1	863	6	PCR-US992-02050-11


```
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-Fls
; US-08-232-463-14

Query Match          3.7%; Score 62.6; DB 1; Length 7218;
Best Local Similarity 2.8%; Pred. No. 2 le-07;
Matches 11; Conservative 231; Mismatches 145; Indels 0; Gaps 0;

Oy 1218 tgggtcgatcaaggaggaatgcgaagatgcgaagatggtggaagagtggtgaga 1277
    |||| |:: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1440 TGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1381

Oy 1278 gtcaagtggaatgcgaagcgcggtgtgtgataggcggtgaattaaagtgctctcg 1337
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321

Oy 1338 gaggttataagagtcgagagtggtgagagaagcgcaatgatgtggaaggttggctaaa 1397
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261

Oy 1398 gaagctatgatgaagaacgtgatcatcaatgaacaattctgaagaatttactacagg 1457
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201

Oy 1458 attattatgaataagcctcaatgaagtgactatatgtattattgtttatggaagc 1517
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141

Oy 1518 tcgaattaagtagttaaagtatatgtattataggaagcacaacgggtcacccc 1577
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081

Oy 1578 ggcagggccaggggtgaaagccgcg 1604
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1080 RRRRRRRRRRRRRRRATCGCAAGCTCC 1054

; Sequence 1, Application US/08522421
; Patent No. 5908973
; GENERAL INFORMATION:
; APPLICANT: Abu-Bakar, Umi Kalsom
; APPLICANT: Barton, Sarah Louise
; APPLICANT: Gallego-Veigas, Pedro Pablo
; APPLICANT: Gray, Julie Elizabeth
; APPLICANT: Grierson, Donald
; APPLICANT: Lowe, Alexandra Louise
; APPLICANT: Picton, Steve
; APPLICANT: Whotton, Lee Colin
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS
; TITLE OF INVENTION: DERIVED THEREFROM
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/522,421
; FILING DATE: 11-JAN-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305868.3
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305869.1
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305859.2
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305865.9
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305866.7
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305867.5
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305860.0
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9305862.6
; FILING DATE: 22-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9314351.9
; FILING DATE: 12-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9320988.0
; FILING DATE: 12-OCT-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: ERT1B
; US-08-522-421-1
```

RESULT 4

US-08-522-421-1

Query Match

2.9%; Score 48.6; DB 3; Length 1669;

```

Best Local Similarity 51.1%; Pred. No. 0.00065;
Matches 114; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1083 gggaaatcgtgacatggtgtcacaattggtggtctcgtcgcataagtcggtggatgc 1142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 960 GGACGAATAGTGAATGGCCACCACAAACAGGTGCTGCACATCCCGGTAGCAGG 1019

QY 1143 tctgtgacatcgtggtgaaattcgtctatcagagcctggtggtggtccggtg 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1020 TTTTTCATCATGTTGGTGGTGAATCTACGCTTGAAGATATATGTAAGAGTCCCTATG 1079

QY 1203 gbtgtcttctcctcaatggttcgatcgaagggtactaatcgaagatgatcgaagatgtgtg 1262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1080 GTGTGAGGCCCATTTCTAGCAGACCAACTGTGTGAACGCAAGGTATCTGAGCCAAATATAC 1139

QY 1263 agagtggtgagagtcagagtcgaatgaagagggcggttg 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1140 AGGTTGGTTCGAATGGAGTTATCGAGAGACGGTCATAG 1182

```

```

RESULT 5
US-08-797-226-1
; Sequence 1, Application US/08797226
; Patent No. 5959180
; GENERAL INFORMATION:
; APPLICANT: MOEHS, CHARLES P
; APPLICANT: ALLEN, PAUL V
; APPLICANT: ROCKHOLD, DAVID R
; APPLICANT: STAPLETON, ANDREW
; APPLICANT: GARBARINO, JOAN E
; APPLICANT: FRIEDMAN, MENDEL
; APPLICANT: BELKNAP, WILLIAM R
; TITLE OF INVENTION: DNA SEQUENCES ENCODING SOLANIDINE
; TITLE OF INVENTION: UDP-GLUCOSE GLUCOSYLTRANSFERASE AND USE TO REDUCE
; TITLE OF INVENTION: GLYCOKALOIDS IN SOLANACEOUS PLANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NANCY J. PARSONS
; STREET: 800 BUCHANAN ST.
; CITY: ALBANY
; STATE: CA
; COUNTRY: USA
; ZIP: 94710
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/797,226
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PARSONS, NANCY J
; REGISTRATION NUMBER: 40,364
; REFERENCE/DOCKET NUMBER: 0011.97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 559-5731
; TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum
; STRAIN: cv. Lemhi Russet
; INDIVIDUAL ISOLATE: SGT1750
; DEVELOPMENTAL STAGE: mature, somatic

```

```

TISSUE TYPE: tuber
IMMEDIATE SOURCE:
LIBRARY: lambda gtl1 cDNA library
CLONE: SGT 1750
FEATURE:
NAME/KEY: CDS
LOCATION: 20..1486
OTHER INFORMATION: /product= "solanidine
OTHER INFORMATION: glucosyltransferase"
PUBLICATION INFORMATION:
AUTHORS: Moehs, Charles P.
AUTHORS: Friedman, Mendel
AUTHORS: Belknap, William R.
TITLE: Cloning and expression of solanidine
TITLE: UDP-glucose glucosyltransferase from potato
JOURNAL: The Plant Journal
VOLUME: 11
ISSUE: 2
DATE: 1997
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 488
US-08-797-226-1

Query Match 2.7%; Score 45.4; DB 4; Length 1607;
Best Local Similarity 47.1%; Pred. No. 0.0047;
Matches 200; Conservative 0; Mismatches 201; Indels 24; Gaps 1;

QY 837 aaagcgcgtgaaactacatgactgactggtctaaactcaaacccgaatcatcggtcggtttac 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 833 AACACAATGAGATTGTTATAGATTGGTTGAATGCACAGAAACCTAANTCGGTCTCTAT 892

QY 897 gttctgttcggagcctcctcgtgaggtcccgaaaccccaaatggaagaataagtagg 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 GTATCTTTTCGGAAGCATGGCTAGATTTCCTGAGAGCCAACTGAATGAATGAATGCCAAGCT 952

QY 957 ctttcagacacaaatcgcagttctctggtgatagaagaagaacagagggcgacgaa 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 953 CTGGATGCTTCAATGTTCTTTTCATTTTCTATTGAGGCTAATGAAGAAACGGCGTCG 1012

QY 1017 caagagcaagcagaagaagaagcgtcgtgagcttcttctgctcgtcaggaactgaa 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1013 TGGTTGCCAGTTGGTAAATTTAGAGGCAAG-----ACTAAA 1048

QY 1077 cgactcgggaaaaatcgtgacatggtgtcacaattggatgttctgacgcataagtcggtg 1136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1049 AAGGGTTTGTACATCAAAAGGGTGGTCCACAGCTTACGATCATGGAACATTCAGCAACA 1108

QY 1137 ggtgctcgtgacgcattcggttggaattctcgtatcagagcctggtggtgtg 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1109 GCGGGTTTCATGACTCATTTGGTGACTAAATTCGGTTCTTGGAAAGCCATCATCTTTTGGCGTG 1168

QY 1197 cccgtggtgtgttctcctcaatggttcgatcaagggaactaatgcgaagatgcgaagat 1256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1169 CCAATGATAACATGCCACTTTTATGCTGATCAATCTACAACGAGAGAGGTAGTCGAGGTT 1228

QY 1257 gttgtg 1261
Db 1229 AGGGG 1233

```

```

RESULT 6
US-09-106-464-1
; Sequence 1, Application US/09106464
; Patent No. 601145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Ghanges, Gurdev S.
; APPLICANT: Kuai, Jian-ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc:Patty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2

```



```
RESULT 9
US-08-188-582-1
: Sequence 1, Application US/08188582
: Patent No. 5534410
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comai, Lucio
: APPLICANT: Dynlacht, Brian D.
: APPLICANT: Hoyer, Timothy
: APPLICANT: Ruppert, Siegfried
: APPLICANT: Tanese, Naoko
: APPLICANT: Wang, Edith
: APPLICANT: Weinzierl, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/188,582
: FILING DATE: 28-JAN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A.
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4615 base pairs
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 538..3300
US-08-188-582-1

Query Match 2.3%; Score 38.8; DB 1; Length 4615;
Best Local Similarity 74.2%; Pred. No. 0.47;
Matches 49; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1606 atgatttaaaatatttttaaaatatttttctactatttaaaactaaaaa 1665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4520 ATAAGTGAATATATATATATTATTATTATTATTATTATTATTATTATTAAAG 4579

Qy 1666 aaaaaa 1671
|||||
Db 4580 AAAAAA 4585

RESULT 9
US-08-646-715-1
: Sequence 1, Application US/08646715
: Patent No. 5637686
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comai, Lucio
```

```
: APPLICANT: Dynlacht, Brian D.
: APPLICANT: Hoyer, Timothy
: APPLICANT: Ruppert, Siegfried
: APPLICANT: Tanese, Naoko
: APPLICANT: Wang, Edith
: APPLICANT: Weinzierl, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/646,715
: FILING DATE: 09-MAY-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/188,582
: FILING DATE: 28-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A.
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4615 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 538..3300
US-08-646-715-1

Query Match 2.3%; Score 38.8; DB 1; Length 4615;
Best Local Similarity 74.2%; Pred. No. 0.47;
Matches 49; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1606 atgatttaaaatatttttaaaatatttttctactatttaaaactaaaaa 1665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4520 ATAAGTGAATATATATATATTATTATTATTATTATTATTATTATTATTAAAG 4579

Qy 1666 aaaaaa 1671
|||||
Db 4580 AAAAAA 4585

RESULT 10
US-08-379-482A-2
: Sequence 2, Application US/08379482A
: Patent No. 5859334
: GENERAL INFORMATION:
: APPLICANT: Brugliera, Filippa
: APPLICANT: Holton, Timothy A.
: TITLE OF INVENTION: GENETIC SEQUENCES ENCODING
: TITLE OF INVENTION: GLYCOSYLTRANSFERASE ENZYMES AND USES THEREFOR
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
```

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10172-0194
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,965C
FILING DATE: August 12, 1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 322745/95
APPLICATION NUMBER: PCT/JP96/03630
FILING DATE: 12-NO. 5936078-1995
FILING DATE: 12-Dec-1996
ATTORNEY/AGENT INFORMATION:
NAME: Lawrence S. Perry
REGISTRATION NUMBER: 31865
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-2400
TELEFAX: 212-758-2982
TELEX: 236262
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1858 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: human
IMMEDIATE SOURCE:
CLONE: F2232
FEATURE:
NAME/KEY: CDS
LOCATION: 26 to 1255
IDENTIFICATION METHOD: by experiment
US-08-909-965C-11

```

Query Match      2.28; Score 37; DB 4; Length 1858;
Best Local Similarity 71.0%; Pred. No. 0.94;
Matches 49; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1603 gccatgattaaaatatattttaaaaataaatattttctactattaaactaaaaaaa 1662
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1763 GCCATGGTCACCTATTATTGGTCAGATAAATCTTTAAATAAAAAAAAAAAAAAAA 1822

Qy 1663 aaaaaaaa 1671
      ||| | | | | | | |
Db 1823 AAAAAAAAAA 1831

```

```

RESULT 12
US-08-658-136-2/c
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:17:46 ; Search time 236.66 Seconds
(without alignments)
1766.547 Million cell updates/sec

Title: US-09-147-955-7
Perfect score: 1671
Sequence: 1 aacacataaaaaaaaaataa.....ctaaaaaaaaaaaaaaaa 1671

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1670.2	100.0	1671	1 X02829	WO9905287 Seq ID 4
2	484.4	29.0	2062	1 X02828	WO9905287 Seq ID 3
3	446.8	26.7	1474	1 X02827	WO9905287 Seq ID 2
4	434.8	26.0	1506	1 X02826	WO9905287 Seq ID 1
5	289	17.3	2105	1 X02831	WO9905287 Seq ID 6
6	103.6	6.2	568	1 T66174	UDP-glucose:thiohy
7	102	6.1	1513	1 T66166	UDP-glucose:thiohy
8	101.8	6.1	1731	1 T10106	Maize UDP-glucose
9	100.4	6.0	566	1 T66173	UDP-glucose:thiohy
10	83.8	5.0	432	1 T66893	Strawberry UDP-glu
11	78.4	4.7	1578	1 T62123	Solanum melongena
12	57.8	3.5	1597	1 V23108	UDP-glucose:flavon
13	48.4	2.9	1650	1 Q74684	Early Ripening Tom
14	45.4	2.7	1607	1 V49609	Potato solanidine
15	45.4	2.7	1624	1 V17054	Glucosyl transfera
16	45.4	2.7	1626	1 X24873	Medium-chain UDP-g
17	44	2.6	794	1 T66179	UDP-glucose:thiohy
18	42.8	2.6	940	1 T66178	UDP-glucose:thiohy
19	40.6	2.4	9789	1 T42852	cDNA encoding Plas
20	40	2.4	405	1 V23952	Plant CGT enzyme D
21	38.8	2.3	1486	1 X02855	Human zslg46 DNA.
22	38.8	2.3	4615	1 Q70724	TATA-binding prote
23	38.8	2.3	4615	1 T42210	Drosophila TATA-bi
24	38.8	2.3	4615	1 T79604	TATA-binding prote
25	38.4	2.3	1738	1 Q56241	Sequence encoding
26	38.2	2.3	2696	1 N60909	Plasmid PAU157 ins
27	38.2	2.3	2696	1 N60889	Plasmid PAU157 seq
28	38.2	2.3	2697	1 N60864	Sequence of plasm
29	38.2	2.3	2966	1 X32291	M. grisea PTH2 gen
30	38.2	2.3	3383	1 T85161	Human chemokine re
31	38	2.3	9502	1 Q74770	Hepatitis C virus
32	37.8	2.3	223	1 T66184	UDP-glucose:thiohy
33	37.8	2.3	561	1 X20491	Human secreted pro

c 34 37.8 2.3 110000 1 V21209_10 Continuation (11 o
c 35 37.8 2.3 110000 1 V21209_11 Continuation (12 o
c 36 37.6 2.3 2401 1 X20301 Borrelia burgdorfe
c 37 37.4 2.2 2418 1 Q27886 P.falciiparum GBP13
c 38 37.4 2.2 4696 1 X03041 Human IL-1ra BAC c
c 39 37.4 2.2 5760 1 N50530 Sequence encoding
c 40 37.4 2.2 6727 1 X02993 Human IL-1ra BAC c
c 41 37.2 2.2 1302 1 X20437 Human secreted pro
c 42 37 2.2 1858 1 T72175 Alzheimer's diseas
c 43 37 2.2 53526 1 T94101 Human PKD1 gene. H
c 44 37 2.2 53577 1 T18551 Human polycystic k
c 45 37 2.2 53577 1 T94108 Human PKD1 locus b

ALIGNMENTS

RESULT 1

X02829
ID X02829 standard; cDNA; 1671 BP.
AC X02829;
DT 14-MAY-1999 (first entry)
DE WO9905287 Seq ID 4.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Torenla hybrira.
FH Key Location/Qualifiers
FT CDS /tag= a
FT FT /product= "protein with flavonoid 5-transglycosylation activity"
PN WO9905287-A1.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNRY) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI: 99-142940/12.
DR P-PSDB; W92950.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PS used to transform plants for improvement of plant coloration
CC This invention describes the isolation of plant proteins which have
CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC in the production of plant varieties with improved colour, for example in
CC the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 1671 BP; 530 A; 344 C; 399 G; 396 T;

Query Match 100.0%; Score 1670.2; DB 1; Length 1671;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacacataaaaaaaaaataaagaagaataattaaaaaaatgggttaacaaagcc 60
DB 1 AACACATATAAAAAAAAAATAAAGAGGAANTATTAATAAAAAAAAAATGGTTACAAACGCC 60
QY 61 attattctactagcaacattccacagcacaagccacataaacacctctctcgagttcgcca 120
DB 61 ATATTCTACTAGCAACATTCCTCCAGCACAAAGGCCACATAAACCTTCTCTCGAGTTGCGCA 120
QY 121 aaagctctcaacacgagatcgtcgaccaagtccattcttcacagtgatcacgat 180
DB 121 AAAGCTCTCTCAACACCGGATACGTCGACCAAGTCACATTCCTTCACAGGTGATACGAT 180
QY 181 tgagacgatcgctctgaaacccgacgagcagagaatcgattctcggtggtatkcag 240
DB 181 TGAGACGATCGCTTTCGAAACCGATCCGAGCAGCAGACATCGATTTCTCGGTGGCATKTYCAG 240
QY 241 attcttagatgatgctttaaagaagcgcagatggcaaaaactacatgctggagatga 300
DB 241 ATTCTTAGCATGTCGCTTAAAGAAAGCGACGATGCGAAAACTACATGTCTGGAGATGA 300

QY 268 gcagcagatggcaaaactacatgctcgagatgagaaagcgcgaagaaagcccttaaaagg 327
 Db 243 GAGACATGGGAGAGAACTACATGTCGGAGATGAAGAGCAGAGGTATATAAGCCCTTGAGCG 302
 QY 328 acactctattaaagctcaacagatgctgcgaagggaagtgaatttaacaatcgctgagct 387
 Db 303 ATACTCTGCAGCCAAATATGTCGATCAAAAGACGACAAA-----ATCAGCT 350
 QY 388 ttgtgtgactctcatctatttttctgtggcagctgaagtggcgctgaagtcgacgtgc 447
 Db 351 TCGTGTGTACTCCACACCTCTTTGTCATGGCGGCCAAGGTGGCGGTGATTCCTCC 410
 QY 448 cgaagtccctcttttgattgaacggctacggttttcgattgctactatttttacttca 507
 Db 411 GGAGCGCGTACTCTGGATTGAGCCAGCTACGGTGTGGATATATTTACTTTTATTTCA 470
 QY 508 atgggtatccgatgatcatcgatcgaggctcagatcaaaatcccaactgccc---aatcttc 564
 Db 471 ACGGCTATAGCAGCAAAATCGATCGGGTTCGGATGCTATTCATCTGCCGGAGGACTCC 530
 QY 565 cgagctctccaagcaagatctccctctcttctactcccttgagcccgcgagattcc 624
 Db 531 CAGTGTGCGCCAGCGTGATTTACCGCTCTTCTCTTCTCCACGCATGAGAGATTCC 590
 QY 625 gaacctaatgaagaaagttcgacagctcgacaaagcaagaaagcaagaggtcttga 684
 Db 591 GTTCACTGATGAAGGAGAAATTTGGAACCTTTAGAAAGGTGAAGAAAACCTTAAGTCTTGG 650
 QY 685 taacacgcttcagcagcattagaaacgcgaacaaactcaagccatcgacagatgaactaa 744
 Db 651 TGAACAGCTTTGATGCTTGGAGCTGTATCGCTCAAGGCCATTTGATTAAGTACGAGATGA 710
 QY 745 tatccatcgcccatataatcccatcatcgatattctcagatggaacagaccctctatca 804
 Db 711 TTGCATTCGGCGGTGATTCCTCCGC---ATTCTTGGACGGTFAAAGATCTTTCGGACA 767
 QY 805 gcaacaaatcttacggtggagacctcttcagaaagcgcgatgaacttatcagactggc 864
 Db 768 GGTCTTTCGGCGGAGATTTTTCGAGAAAGGTGCAATGACGAGATGCTCGAATGCT 827
 QY 865 taactcaaaaccccaatcatcgctcgcttactcgcttgcgtggagcctcctgagctcc 924
 Db 828 TGAGCAGCAATCTCGATCTCGGTGCTTACGTTTTCGTTTTCGGAAGCTTCGTTAATACGA 887
 QY 925 cgaaccccaaatggagaaatagcaatagggtcttcagacaccacaaatcccgactctct 984
 Db 888 CGAAGTCGCAATGGAAGATAGCAGAGGCTGTAGATTGTTGGAGGCCCTTTTGT 947
 QY 985 ggtgataagaagaacgaagagggcgacgaagcaagagcaagcaagaagaagaagc 1044
 Db 948 GGTGTAAAGATAAAGCAAGAGAGAGAGGTA-----TTGATAAGT 988
 QY 1045 tgcgtgactcttctgactcgacgaactgaacactcgggaataatgtagactgct 1104
 Db 989 TGCATGGAGGAGTTGA-----AACGAGTGGGGAATTTGATCTTGTGT 1034
 QY 1105 cacaattgagttctgacgataagtggtgggagctcttcgacgcatcggttgga 1164
 Db 1035 CTCAATGGAAGTCCTGACGCATCTCTGTTGGGATGTTCTGACACACTCGGGGTGA 1094
 QY 1165 attctgctatcgagagcctggttgggtgctgctggtggtgcttccatggttgc 1224
 Db 1095 ATTCGACTCTAGAGATATATCTTTCGGGTTCCGATGGTGGCTTTTCCGACGTGTTG 1154
 QY 1225 atcaagggactaatgcgaagatgatacgaagatgtagagagtggtgtagagtcagag 1284
 Db 1155 ATCAAGGGACGAATCGAAGCTGTATGGAGGATGTGTGAGACGGGTGTGAGAGTAGAG 1214
 QY 1285 tgaatgagaagggcggttggttgtagcgtgagattaaagatgctgctcgaggtta 1344
 Db 1215 CTAATCAGGAGGGTAGCGTCTGTATGATGAATAGGAGATGATTTGAGGAGGTGA 1274

QY 1345 t-----aaagagtcgagagtgtagaagaagcgaatgatgtgaaaggggttgg 1392
 Db 1275 TGGATGGGGAGAGAAAGATAGTAACTTAGAGAGAGTCTGGCAAGTGGAAAGGATTGG 1334
 QY 1393 ctaaaagaagctatggatgaagaacgtggtatcatcaatgaacaatctgaagaatttatta 1452
 Db 1335 CAAGAAAGCTATGGAGGAAGA---TGGATCTTCAGTTAACAACTTCAAGGTCTTCTTG 1391
 QY 1453 ctaggattat 1462
 Db 1392 ATGAGGTGT 1401

RESULT 3

X02827
 ID X02827 standard; cDNA; 1474 BP.
 AC X02827;
 DE 14-MAY-1999 (first entry)
 KW W0905287 Seq ID 2.
 OS Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
 FH Perilla frutescens.
 FT Key Location/Qualifiers
 CDS 29..1360
 FT /*tag- a
 FT /product- *protein with flavonoid 5-transglycosylation activity*
 FT
 PN W0905287-AL.
 PD 04-FEB-1999.
 PF 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNR) SUNTORY LTD.
 PG Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 PI Yamazaki M,
 DR WPI: 99-142940/12.
 DR P-PDB: W92948.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
 PS This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.
 SQ Sequence 1474 BP; 369 A; 348 C; 438 G; 319 T;

Query Match 26.7%; Score 446.8; DB 1; Length 1474;
 Best Local Similarity 63.0%; Pred. No. 3.8e-101;
 Matches 896; Conservative 2; Mismatches 434; Indels 90; Gaps 10;

QY 57 cgcatattctactagcaacattccacagacaaagccacataaaccttctctcgagttc 116
 Db 38 CGCGCGTGTCTGTAGCAACGGTTCCGGCGAAGGCCACATAAATCCCGCCTCCAATTC 97
 QY 117 qcacaaaggtctctcaaccccgatacgtcgaccagtcacatctctcagagtgatc 176
 Db 98 GCCAAGAGACTCTTAAAGCCGG-----CACTGACGTACGTTTTTCACGAGGTTTAT 151
 QY 177 gcattgagacgctgcgtctgaaacccgatccgagcagcagaa-----tcgat 224
 Db 152 GCATGGCGCGCATGTCACCAACAGAGCTCCGCGCTCGCGGAAACCCACCGCGGCTCGAC 211
 QY 225 ttcgtggcatktycagattctcagatgctttaaagaagcgacagtggaacaaac 284
 Db 212 TTCGTGGCGTTCCTCCACGGCTACGACGACGGGCTGAAGCCCGGGCGGGAAGCGC 271
 QY 285 tacatgtcggagatgagaagcgcggaacgaaggccttaaaggacactcttattaagctc 344
 Db 272 TACATGTCCGAGATGAAGACCGCGGCTCCGAGGCTTTAAGAAACCTCTTCTCAACAAC 331
 QY 345 aacgatgctcgatgggaagtgaattacaatcgcgagcttgggtgactctcat 404
 Db 332 GACGA-----CGTACTTTCGTCGTCTACTCCAC 361

```

RESULT      4
X02826      X02826 standard; cDNA: 1506 BP.
ID          AC X02826;
DE          14-MAY-1999 (first entry)
DF          WO9905287 Seq ID 1.
DE          Plant: flavonoid 5-transglycosylation activity; STG; variety; colour; ds.
KW          Parilla frutescens.
OS          Location/Qualifiers
FH          Key 17..1399
FT          CDS
FT          /tag= a
FT          /product= "protein with flavonoid 5-transglycosylation
FT          activity"
PN          WO9905287-A1.
PD          04-FEB-1999.
PF          16-JUL-1998; J03199.
PR          25-JUL-1997; JP-200571.
PA          (SUNR ) SUNTORY LTD.
PI          Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
EI          Yamazaki M.
DR          WPI; 99-142940/12.
DR          P-PSDB: W92947.
PT          Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT          used to transform plants for improvement of plant coloration
PS          Disclosure; page 49-52; 89pp; Japanese.
CC          This invention describes the isolation of plant proteins which have
CC          flavonoid 5-transglycosylation (5NG) activity. Such proteins can be used
CC          in the production of plant varieties with improved colour, for example in
CC          the production of flower varieties with desired coloration and cut
CC          flowers harvested from them.
CC          Sequence 1506 BP; 364 A; 351 C; 449 G; 342 T;
SQ          Sequence 1506 BP; 364 A; 351 C; 449 G; 342 T;

Query Match 26.0%; Score 434.8; DB 1; Length 1506;
Best Local Similarity 62.6%; Pred. No. 3.5e-98;
Matches 890; Conservative 2; Mismatches 434; Indels 96; Gaps 10

Qy 57 cgcatactctatagcaacattccagcacaaagccacataaaacctctctcaggttc 116
Db      |||||
Db 26 CGCGCGGTGCTGTAGCAACGGTTTCGTGGCAAGGCCACATAAATTCGGCGCTCCAAATTC 85
Qy 117 gccaaaaggctctcacaacccggtacgtcgaccgaagtcacattcttcacgagtgatac 176
Db      |||||
Db 86 GCCAAGAGACTCTAAAGCGGG-----CACTGAGCTCACATTTTTCAGGAGGTTTAT 139
Qy 177 gcattgagacgatgcgtctcgaaacccgatacgcgacgacgaa-----tcgat 224
Db      |||||
Db 140 GCATGGCGCGCATGGCCACACAGCCTCCGCGCTGGCGGAACCCACCACCGGGCTCGAC 199
Qy 225 ttctgtgcatktycagattctcagatgatgctttaaagaagcgacgatggcaaaac 284
Db      |||||
Db 200 TTCGTGGCGTTCTCCGACGGCTACGACGACGGGCTGAAGCCCTGGCGGCGAGGAGCGC 259
Qy 285 tacatgtcggagatgagaagcgggaacgaaggaccttaaaggacactcttattaaotc 344
Db      |||||
Db 260 TACATGTCGAGATGAAGACCCCGGCTCCGAGGGCTTAAAGAAAGCTCTCTTCTCAACAA 319
Qy 345 aacgatctcgatgggaagtgaatttacaatcgcgtgagctttgttggtatctctcat 404
Db      |||||
Db 320 CACGA-----CGTACGTTTGGTGGTCTACTCCAC 349
Qy 405 ctattttctgggcagctgaagtggcgctgaagtcgacgtgcgagtgccctctcttgg 464
Db      |||||
Db 350 CTCCTTTGCATGGGCGGAGGTGGCGGTGAGTCCAGGTCCCGAGGCCCTTCTCTCTGG 409
Qy 465 atgaacgcgtcaaggttttcgatgtgtactatttttacttcaatgggtatgcgagat 524
Db      |||||
Db 410 GTCGAGGCCCGCCACCGTGTGTCATATATTACTTCTACTTACGGGTACGACAGCAG 469
Qy 525 atcgtatcgggctcagatcaaatcccaactgcgcatacttccgcagctctccaagaagat 584
Db      |||||
Db 470 ATGCAGCCGGTTTCCGACGAATATCAGTTCCTCGGCTTCCACCGCTCGAGCGCGAGT 529

```

/product= "protein with flavonoid 5-transglycosylation activity"

W090905287-Al.
PN
PD 04-FEB-1999.
PF 16-JUL-1996; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTOBY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M.
DR WPI; 99-142940/12.
P-PSDB; W92952.

Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration

PS Disclosure; Page 67-71; 89pp; Japanese.

This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.

CC
CC
CC
CC
Sequence 2105 BP; 378 C; 429 G; 629 T;

Query Match 17.3%; Score 289; DB 1; Length 2105;
Best Local Similarity 56.1%; Pred. No. 3.6e-62;
Matches 847; Conservative 2; Mismatches 557; Indels 105; Gaps 12;

Qy 36 aaaaaaaatgggttaacaaagccatatctactagcaacattcccagcacaaagcacc 95
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AAGCAAGAATAATGGTCAGCCTCATGTCATTAAACAACATTTCCAGCACANGGCCAT 388

Qy 96 ataaaccttctcgagtgcgcaaaaggctctctcaaccggatacgtcgaccaagtc 155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 ATTAATCCAGCACTTCAATTTGCCAAGAATCTGTCAAGATGGGCATAG-----AAGTG 442

Qy 156 acattttccagagtgatagcagcatgagacgatgcgctcgaaaccgatccgagc--- 212
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 ACATTTTCACAAAGCATTTATGCCCAAAGCGGTATGGATGAANAATTCATTAAATGCA 502

Qy 213 ---agcagaatcgatttgtggcatktycagatcttLacgatgatgctttaaagaagc 269
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 CAAAAGGATGAATTTTCATTCCATTTCCGATGSCITTTGTAAGSTTTTGATCATTTCA 562

Qy 270 gacgatgccaaaactacatgtcggagatgaaagcgcgaagcgaagcgtttaagagac 329
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 AAAGACCCTGTATTTTACATGTGCACACTTCGTAATGTGGAATGGAATGTGCAAAAA 622

Qy 330 actcttattaagctcaacgatgctgcgtagggagtgaaatgttacaatgcgtagcttt 389
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 ATAATCTCATTTGCTC-----TGAAAATGSACAGCCTATAACTTGC 664

Qy 390 gtggtgtactctcatctatttctgtggcagctgaagtgcgctgaagtcgacgtgcog 449
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 CTACTTTACTTCCATTTTCTCTGTGGCAGCAGAGGTAGCAGTGAAGTTTCATTCCT 724

Qy 450 aglgcctctcttgattgaacgcgctacggttttcgatgtgtactatttttactccaat 509
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 TCTGCTCTCTTTGGAGTCAACACAGCAACAATATTGGACATATATTACTTCAACTTTCAT 784

Qy 510 gggtatgcgatgatagatgcggtgcgaatca-----aatcaactgcgc 557
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 GGATATGA AAAAGCTATGGCTTAATGAATCCAATGATCCAAATTTGGTCCATTCAACTTCC 844

Qy 558 aa tttccgcagctcttcaaagaagatctcccccttcttccatccctt----- 606
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 GGGCTTCCACTACTGGMAACTCGAGATCTCTCTTCAITTTTTTACTTCTTATGGTGCAAA 904

Qy 607 -cgagccgcgcgagattccgaaccttaatgaagaaagttcgacacgctcgacaaagaa 665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 GGGAGTCTTTGAGATTTGCACTTCCACCAATTCAAAGAAATTTGATAGACACATTAGATGCTGAA 964

Qy 966 ccgaagcgaaaggtctttagtaaacacgcttcgaacgatagaaacgcaaacactcaagac 725
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 ACCATCTCTTACATCTCTGATACATCTTCTGATCAATTTAGCCCTCTGCGCTCTCATCTCA 1024


```
FT FT acid change from Ala-2 to Val-2"
FT FT /tag= c
FT FT /note= "G is A in clone pGL6-14; there is no amino
FT FT acid change"
FT FT 99. .99
FT FT /tag= d
FT FT /note= "codon AAG (Lys) is present between clone
FT FT pGL9 bases 98 and 99 in clone pGL6-14"
FT FT 195
FT FT /tag= e
FT FT /note= "G is C in clones pGL6-14 and pGL3-22,
FT FT causing an amino acid change from Val-43
FT FT to Leu-43"
FT FT 245
FT FT /tag= f
FT FT /note= "T is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 292
FT FT /tag= g
FT FT /note= "T is C in clone pGL3-22, causing an amino
FT FT acid change from Leu-75 to Pro-75"
FT FT 323
FT FT /tag= h
FT FT /note= "A is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 331
FT FT /tag= i
FT FT /note= "A is G in clone pGL3-22, causing an amino
FT FT acid change from Glu-88 to Gly-88"
FT FT 345
FT FT /tag= j
FT FT /note= "A is C in clone pGL3-22, causing an amino
FT FT acid change from Asn-93 to His-93"
FT FT 354
FT FT /tag= k
FT FT /note= "G is C in clone pGL3-22, causing an amino
FT FT acid change from Glu-96 to Gln-96"
FT FT 368
FT FT /tag= l
FT FT /note= "A is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 377
FT FT /tag= m
FT FT /note= "A is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 398
FT FT /tag= n
FT FT /note= "A is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 410
FT FT /tag= o
FT FT /note= "A is T in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 458
FT FT /tag= p
FT FT /note= "T is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 465
FT FT /tag= q
FT FT /note= "A is C in clone pGL3-22, causing an amino
FT FT acid change from Ile-133 to Leu-133"
FT FT 473
FT FT /tag= r
FT FT /note= "T is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 512
FT FT /tag= s
FT FT /note= "T is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 515
FT FT /tag= t
FT FT /note= "A is T in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 526
FT FT /tag= u
FT FT /note= "T is C in clone pGL3-22, causing an amino
FT FT acid change from Ala-153 to Val-153"
FT FT 568
FT FT /tag= v
FT FT /note= "C is T in clone pGL3-22, causing an amino
FT FT acid change from Pro-167 to Leu-167"
FT FT 587
FT FT /tag= w
FT FT /note= "A is G in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 623
FT FT /tag= x
FT FT /note= "C is G in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 647
FT FT /tag= y
FT FT /note= "G is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 678. .679
FT FT /tag= z
FT FT /note= "bases 678-679 are AT in clone pGL3-22,
FT FT causing an amino acid change from Arg-204
FT FT to Ile-204"
FT FT 685
FT FT /tag= aa
FT FT /note= "C is T in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 714
FT FT /tag= ab
FT FT /note= "A is G in clones pGL3-22 and pGL4-2,
FT FT causing an amino acid change from Ser-216
FT FT to Gly-216"
FT FT 759
FT FT /tag= ac
FT FT /note= "G is A in clone pGL3-22, causing an amino
FT FT acid change from Ala-232 to Thr-232"
FT FT 761
FT FT /tag= ad
FT FT /note= "G is C in clone pGL3-22, causing an amino
FT FT acid change from Ala-232 to Thr-232"
FT FT 766
FT FT /tag= ae
FT FT /note= "G is A in clones pGL3-22 and pGL4-2,
FT FT causing an amino acid change from
FT FT Arg-234 to Lys-234"
FT FT 794
FT FT /tag= af
FT FT /note= "A is C in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 809
FT FT /tag= ag
FT FT /note= "C is T in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 814
FT FT /tag= ah
FT FT /note= "G is C in clone pGL3-22, causing an amino
FT FT acid change from Gly-249 to Ala-249"
FT FT 848
FT FT /tag= ai
FT FT /note= "C is T in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 901
FT FT /tag= aj
FT FT /note= "C is T in clone pGL3-22; there is no amino
FT FT acid change"
FT FT 936
FT FT /tag= ak
FT FT /note= "G is A in clone pGL3-22, causing an amino
FT FT acid change from Gly-290 to Arg-290"
FT FT 959
```

```

FT      /*tag= al
FT      /note= "C is A in clone pGL3-22; there is no amino
FT      acid change"
FT      968
FT      conflict
FT
FT      /*tag= am
FT      /note= "G is A in clone pGL3-22; there is no amino
FT      acid change"
FT      973
FT      conflict
FT
FT      /*tag= an
FT      /note= "A is C in clone pGL3-22, causing an amino
FT      acid change from Lys-302 to Thr-302"
FT      1013
FT      conflict
FT
FT      /*tag= ao
FT      /note= "C is T in clone pGL3-22; there is no amino
FT      acid change"
FT      1117
FT      conflict
FT
FT      /*tag= ap
FT      /note= "G is T in clones pGL2-7 and pGL2-25,
FT      causing an amino acid change from
FT      Val-350 to glu-350"
FT      1121
FT      conflict
FT
FT      /*tag= aq
FT      /note= "G is T in clone pGL2-7; there is no amino
FT      acid change"
FT      1130
FT      conflict
FT
FT      /*tag= ar
FT      /note= "C is T in clone pGL2-7; there is no amino
FT      acid change"
FT      1161
FT      conflict
FT
FT      /*tag= as
FT      /note= "T is C in clone pGL2-7; there is no amino
FT      acid change"
FT      1169
FT      conflict
FT
FT      /*tag= at
FT      /note= "A is G in clone pGL2-7; there is no amino
FT      acid change"
FT      1193
FT      conflict
FT
FT      /*tag= au
FT      /note= "T is G in clone pGL2-7; there is no amino
FT      acid change"
FT      1202
FT      conflict
FT
Query Match
Best Local Similarity 6.18; Score 102; DB 1; Length 1513;
Matches 235; Conservative 0; Mismatches 190; Indels 9; Gaps 1;
QY 1090 tcgtgacatggtgcacaaattgagttcttcgacgataagtcggtggatgcttcgtga 1149
DB 1078 TTGTTCTTGTGTAAACAGCTTGAGCTTTTACGTCATGATGATAGCTTCTTTTGA 1137
QY 1150 cgcattgcggttgaattctgctatcagagccgtgcttgcgtggtgcctgcgtgct 1209
DB 1138 CTCACCTGCGGTGGAACCTGACGCTTGAAGGATGAGTTGGAGTTCGCGATGGTGTG 1197
QY 1210 ttccctcaatggttcgatacgaaggactaatcgaagatgatcgaagatggtgagagtg 1269
DB 1198 TGCCCTCAGTGGAGTGATCAGATGAATCATCTAAGTTTGGAGGAGGTTTGGAGAGTTG 1257
QY 1270 gctgagagtcagagtgatgagaaaggcgcgttcttgcagagcgtgagattgaagagt 1329
DB 1258 GGTATAGACGAAGAGGAAAGCTGGGGGAGGAGTTGTGAAGACGATGAGTGGTGAGGT 1317
QY 1330 -----gctctcgaggttataaagagtcgagttgagaaagcgaatgatgt 1380
DB 1318 GTTTCAGAGGAGTATGCAAGCAGAGAGTAGTGTGGAGATAGAGAGAGTTCTAAGAGT 1377
QY 1381 ggaagggtttgctaagaagcattgatgaagaacgtggatcataatgaacaatctga 1440
DB 1378 GGAAGAGATTGGCTGTGAAGCGCATGAGTGAAGGAAAGAGCTGTGATCGGAGCATTAATG 1437
QY 1441 agaatcttactaggattattataaataatgcctcataagttgtactatatattat 1500
DB 1438 AGTTTGTGAGAGCTAGGGGAAGAACATTGAGAGGTAAATGAGATTGTGTAATCTTGTGT 1497

```

```

QY 1501 tattgttttatgg 1514
DB 1498 GTTGTGTGTGTG 1511

RESULT 8
T10106
ID T10106 standard; cDNA to mRNA; 1731 BP.
AC 27-OCT-1996 (first entry)
DE Maize UDP-glucose indol-3-yl:acetyl:glucosyl transferase sequence.
KW UDP-glucose-indol-3-yl:acetyl:glucosyltransferase; enzyme;
KW transgenic plant; EC-2.4.1.131; crop improvement; corn; cereal;
KW grass; IAGLU-transferase; ds.
OS Zea mays.
FH Key Location/Qualifiers
FT cds 57..1472 /*tag= a
FT misc_binding 858 /*tag= b
FT FT /*note= "start of UDP binding region"
FT W09600291-A1.
PD 04-JAN-1996.
PR 20-JUN-1995; U07820.
PR 24-JUN-1994; US-265427.
PA (RESE) RESEARCH CORP TECHNOLOGIES INC.
PI Banduraki RS, Szczylowski K, Szerszen JB;
DR WPI; 96-068875/07.
DR P-PSDB; R99999.
PT UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein
PT - used in sense or anti-sense orientation for the genetic control of
PT plant rAA levels and plant growth
PT Claim 2; Page 21-23; 41pp; English.
CC This nucleotide sequence encodes a UDP-glucose-indol-3-
CC ylacetylglucosyltransferase (IAGLU-transferase) from maize (Zea mays).
CC Over-expression of this sequence in transgenic plants will result
CC in at least 1 of the following effects: inhibition of apical
CC dominance; inhibition of stem elongation; inhibition of cell
CC enlargement and increased numbers of stems per plant as compared
CC with a wild-type plant. The ability to control the expression of
CC this enzyme allows the control of free indoleacetic acid levels in
CC plants, thereby effecting plant growth rates. Antisense constructs
CC may be used to inhibit synthesis of this enzyme.
CC Sequence 1731 BP; 298 A; 528 C; 601 G; 304 T;
SQ

Query Match
Best Local Similarity 6.18; Score 101.8; DB 1; Length 1731;
Matches 278; Conservative 0; Mismatches 232; Indels 27; Gaps 1;
QY 822 ggagaccttctcagaaaagccgatgaaacttcaatgactggtctaaactcaaaacccgaa 881
DB 828 GGGGCAACCTGGTGAAGCGGAGGATCGGTGCACCAAGTGGCTAGACACCAAGCCGAC 887
QY 882 tcattcggtcgtttacttcgttcggagcctctcaggtcccgaaaccccaaatcgaa 941
DB 888 CGCTCGTGGGCTAGCTCTCCTTCGGAGCGCTCGCTCCCTGGGCAACGCCAGAGAGAG 947
QY 942 gaaatagcaatagggttcttcagacacacaaatcgccagttctctgggtgataagaagaac 1001
DB 948 GAGCTCGCGCGCGGCTCTCTCGCGCGGCAAGCCGTTCCTGTGGTGTGAGGGCCAGC 1007
QY 1002 gaagagggcgacgaacaagagcagaagcagaagcagctcgtcgtctctttgt 1061
DB 1008 GACGAGCACCAGTCCCGCGCTATCTCTGGCCGAGCGGACGCGGCGCGC----- 1060
QY 1062 cgtcaggaactgaacgactcgggaaatcgtgacatggtgctcacaattgagatgtctg 1121
DB 1061 -----CGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1100
QY 1122 acgataagtcggtggagatcctcgtgacccattcgggttggaattctctatcagagac 1181

```

```

Db 1101 CGGCACCGCGCGCTGGCGTCTTCGTCACCCACTGCGGTTGGAACTCCACGCTGGAGCG 1160
Qy 1182 ctggtctgtgtgcccgtggtgtgtcttctcaatggttcgatcaaggagactaatgcy 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1161 CTCAGCTTCGCGCTGCTATGTTGGCGATGGCGCTGTGGACGACCGACCGCAACGCT 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1242 aagatgatcaagaatgctggtgagagagtgctggaaglcacagtgcaatgaagagcgac 1301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1221 CGGACGCTCGAGCTCGCTTGGCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCG 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1302 gttgttgataggcgtgagatgaagggtgcgtctcggagggtataaagagtcgagag 1358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1281 GTGTTCTTCGCGGGAAGTGGAGCGTGCCTGCGCGCGCTCATGTGACGGGCGGAG 1337

RESULT 9
T66173
ID T66173 standard; cDNA to mRNA; 566 BP.
AC T66173;
DT 15-JUL-1997 (first entry)
DE UDP-glucose:thiohydroximate S-glucosyltransferase clone pGL2-7.
KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;
KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola; ds.
OS Brassica napus cv. Westar.
PN EP-7711878-A1.
PD 07-MAY-1997.
PF 31-OCT-1995; 402425.
PR 31-OCT-1995; EP-402425.
PA (CAN ) NAT RES COUNCIL CANADA.
PA (PLB ) PLANT GENETIC SYSTEMS NV.
PI Grootwassink JWD, Hemmingsen SM, Kolenovsky AD, Peferoen M;
PI Read DM, Underhill EW, Van Audenhove K;
DR WPI: 97-247418/23.
PT Plants genetically transformed to interfere with
PT UDP-glucose:thiohydroximate S-glucosyltransferase gene expression
PT - useful for production of rapeseed oil with reduced glucosinolate
    content
PS Example 2; Page 18-19; 35pp; English.
CC Brassica napus UDP-glucose:thiohydroximate S-glucosyltransferase
CC (S-GT) partial cDNA clones pGL2-7 (T66173) and pGL2-25 (T66174)
CC were obt'd. by PCR-RACE (see also T66167-72) of B. napus cDNA.
CC Analysis of the sequences revealed an open reading frame of about
CC 470 bp, a 104 3' untranslated region and polyA+ tail for both
CC clones. The amino acid sequence of the encoded protein revealed
CC part of S-GT peptide 2 (W09827), as expected because this sequence
CC was used for PCR-cloning, and the complete S-GT peptides 5, 6 (with
CC one amino acid difference) and 7 (W09830-32). A full-length cDNA
CC clone (T66166) for B. napus S-GT (W09825) was subsequently obt'd.
CC Sequence 566 BP; 160 A; 54 C; 193 G; 159 T;

Query Match 6.0%; Score 100.4; DB 1; Length 566;
Best Local Similarity 53.9%; Pred. No. 7.3e-16;
Matches 234; Conservative 0; Mismatches 191; Indels 9; Gaps 1;

Qy 1090 tcgtgacatggtctcacatgattgtctgacacataaagtcggtggtggtctctga 1149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 TTTTCTTGTGTTAACCACCTTGAGGTTTACGTCATGGATCTATAGTGTGTTTGTGA 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1150 cgcattcggttggaattctgctatcgagagcctggttggtggtgcccgtggtgtct 1209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 CTCACCTCGGTTGGAACTCGACGCTGGAAGGTTGAGTTTGGAGTTCGATGCTGGGTG 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1210 ttctcgaatggttcgatcaaggagactaaagcgaagatgatcgagagatgctggaaggatg 1269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 TGCCGCGAGTGGAGTGATCAGATGAATGATGCTAAAGTTTGTGGAGGAGGTTTCGAGAGTTG 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1270 gttgagagtcagagtgaatgaggaaggcggtgtgtgtgtgtaggagcgtgagattaagaggt 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 GGTATAGAGCGAAAGATGAAGCTGGGGAGGAGTGTGAAGAGCGATCAGGTGCTGAGGT 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1330 -----gcgtctcgagaggttataaagagtcgagaggttagagagaagcgaatgatgt 1380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 303 GTTTGAAGGAGTGATGGAAGGAGAGAGTAGTGTGGAGATTAGAGAAAGTTCTTAAGAAAT 362
Qy 1381 ggaagggtttggctaaagaagctatggaagaacgtggtgacatcaatgaacaatctga 1440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 GGAAGAGTTTGGCTGTGAAGCGATGAGTGAGGAGGAAGCTCTGATCGGAGCATTAATG 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1441 aqaattttactactagattattaatgaataaagccctcataaagtctactatatattat 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 AGTTTGTGAGAGTTTAGGGAAGAACATTTGAGAGGTAATGAGATTGTTAAATCTTGTGT 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1501 tatttgtttatgg 1514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 GTTTGTTGTTGTTG 496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
T68693
ID T68693 standard; cDNA; 432 BP.
AC T68693;
DT 05-JAN-1998 (first entry)
DE Strawberry UDP-glucuronosyl transferase cDNA.
KW Strawberry; UDP-glucuronosyl transferase; ERT1b; fruit; ripening;
KW storage; gene regulation; ss.
OS Pragmaia ananassa Duch. cv. Brighton.
PN W09721816-A1.
PD 19-JUN-1997.
PF 12-DEC-1996; G03076.
PR 13-DEC-1995; GB-025459.
PA (ZENE ) ZENCA LTD.
PI Manning K;
DR WPI: 97-332787/30.
PT Vectors encoding new enzymes for regulating ripening of fruit - used
PT particularly to improve storage properties, processing
PT characteristics etc, especially in strawberries
PS Claim 1; Page 23-24; 43pp; English.
CC This cDNA sequence comprises a UDP-glucuronosyl transferase (ERT
CC 1b) cDNA clone isolated from a ripe strawberry fruit cDNA library
CC by differential screening. 9 ripening-enhanced clones (see T68687-
CC 95) clones were isolated from the library, and are deposited at
CC NCIMB 40693. Any of these 9 sequences can be used as a regulatory
CC sequence in claimed vectors for genetic transformation of plant
CC cells to regulate fruit ripening. The vectors also comprise a
CC promoter and transcription termination sequence. Also claimed are
CC plants (especially strawberry) and propagating materials containing
CC these vectors. Stable integration of the regulatory sequences, or
CC their complements or genomic equivalents, into a plant is used to
CC modulate fruit ripening by overexpression or downregulation of an
CC endogenous plant gene. Slowing ripening of fruit improves
CC resistance to damage during harvesting, packaging and transport.
CC It also extends shelf life, improves storage properties, processing
CC characteristics, flavour and aroma, and modifies colour and
CC increases resistance to post-harvest fungi.
SQ Sequence 432 BP; 101 A; 84 C; 136 G; 94 T;

```

```

Query Match 5.0%; Score 83.8; DB 1; Length 432;
Best Local Similarity 59.8%; Pred. No. 8.2e-12;
Matches 136; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 1083 gggaaaaactgacatggtgtctcacatggtgtctgacacataaagtcggtggatgc 1142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 GCGAAGTTGTGCAATGGATGGATCCACAGAGAGAGATTTTGGAGCATCCTTCGACGGCTTGC 171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1143 ttctgacgaatcggttggttggaattctgctatcgagagcctggttgggtggtgcccgtg 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 TTTGTGACTCATTCGCGGTGGAATCAACCATGGAGTCACTACCTCAGGAATGCCCGTG 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1203 gtgtgcttctcctcaatggttcgatcaaggaggaactaatgcgaagatgatcgagaatgtgtg 1262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GTGCGATTCACACATGGGTGACCAAGTGACCGACGCCAAGTATTTGTGTCACAGATT 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1263 aggaagtgtgtgagagtcagagtgaatgaggaaggcggtgtgtgtgtgtgtgtgtgt 1310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


CC 5' sequences. GST is involved in the biosynthesis of steroidal
CC glycoalkaloids in solanaceous plants, catalysing the UDP-glucose
CC dependent conversion of the aglycone solanine to gamma-chaconine.
CC The invention relates to DNA sequences which encode SGT, and their
CC use, particularly use of an antisense DNA construct to inhibit SGT
CC activity and glycoalkaloid levels in solanaceous plants.
CC Transgenic plants are claimed, particularly tomato and potato, that
CC have reduced contents of toxic steroidal glycoalkaloids. The
CC inactivation of glycoalkaloid biosynthetic pathways is beneficial
CC to reduce or eliminate glycoalkaloid biosynthesis during shipping
CC and storage. Fragments of SGT nucleic acids are also useful as
CC probes for detection and isolation of SGT-encoding genes.
CC SQ Sequence 1607 BP; 484 A; 306 C; 334 G; 483 T;

```

Query Match          2.7%; Score 45.4; DB 1; Length 1607;
Best Local Similarity 47.1%; Pred. No. 0.038;
Matches 200; Conservative 0; Mismatches 201; Indels 24; Gaps

QY 837 aaagcggatgaactacatggactggctaaactcaaacccgcaatcatcggtgtttac 896
   || | || || | || | || | || | || | || | || | || | || | || | || |
DB 833 AACACAATGAGATTCTTATAGATTGGTTGAATGCACAGAAACCTAATAATCGTTCCTAT 892

QY 897 gtttcgtcggaggacctcctgaggctcccgaaccaccaatggagaataagcaatagg 956
   || | || || | || || | || | || | || | || | || | || | || | || |
DB 893 GTATCTTTCGGAGCATGGCTAGATTCTTGAGCCCACTGAATGAANTAGCCCCAAGCT 952

QY 957 ctttcagacacaaatcgccagtctctcgtgggtgataagaagaacgaaggcgcacgaa 1016
   || | || || | || || | || | || | || | || | || | || | || | || |
DB 953 CTGGATGCTTCAAATGCTCCTTCATTCTTGTATTTGAGGCCCTAATGAAGAAACGCCGTCG 1012

QY 1017 caagaacaagcagaagaagaagaagtgctgagcttcttggatcgtaacgggaactgaa 1076
   || | || || | || || | || | || | || | || | || | || | || | || |
DB 1013 TGGTTGCCAGTGGTAATTATGAGGACAAG-----ACTAAA 1048

QY 1077 cgactcgggaaaatcgtagacatgggtgctcacaaattgggatcttcgacgcataagtcggtg 1136
   || | || || | || || | || | || | || | || | || | || | || | || |
DB 1049 AAGGGTTTGTATCATCAAGAAGGTGGGTCCACACAGCTTAGCATCATGGAACATTCAGCAACA 1108

QY 1137 ggatgcttcgtgacgcattgcggttggaattcgtctcgagagcctggcttgtggtg 1196
   || | || || | || || | || | || | || | || | || | || | || | || |
DB 1109 GCGCGGGTTCATGNACTCATTTGTGGTACTAATTCGGTTCTGGAAGCCATCACACTTTTGGCGTG 1168

QY 1197 ccggtgggtgtgcttccctcaatggttcgatcaagggaactaatggaagatgatcgcaagat 1256
   || | || || | || || | || | || | || | || | || | || | || | || |
DB 1169 CCAATGATACATATGGCCACTTTATGCTGATCAATTTCTACAACGAGAAGGTATGCGAGTT 1228

QY 1257 gtgtg 1261

DB 1229 AGGGG 1233
```

RESULT 15

V17054	
ID	V17054 standard; cDNA; 1624 BP.
AC	V17054;
DT	08-JUN-1998 (first entry)
DE	Glucosyl transferase (Grase) encoding wound inducible gene (TWI1).
KW	Glucosyl transferase; Grase; TWI1; tomato; signalling pathway;
KX	salicylic acid; jasmonic acid; ethylene; wound inducible gene;
OS	plant defence protein; plant response; tobacco; rice; ss.
KS	Lycopersicon sp.
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	2..1410
FT	/tag= a
FT	/product= "glucosyl transferase"
FT	/note= encodes Grase from amino acid position 5
FT	shown in W47172
PN	W09745546-A1.
PD	04-DEC-1997.
PF	30-MAY-1997; G01473.
PR	31-MAY-1996; GB-011420.
PA	(UYYO-) UNIV YORK.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: August 1, 2000, 13:07:11 ; Search time 9137.45 Seconds
(without alignments)
326.374 Million cell updates/sec

Title: US-09-147-955-7
Perfect score: 1671
Sequence: 1 aacacataaaaaaaaaataa.....ctaaaaaaaaaaaaaaaaaa 1671

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
 - 1: gb_bal.*
 - 2: gb_ba2.*
 - 3: gb_om.*
 - 4: gb_ov.*
 - 5: gb_pat.*
 - 6: gb_ph.*
 - 7: gb_pll.*
 - 8: gb_pl2.*
 - 9: gb_pr1.*
 - 10: gb_pr2.*
 - 11: gb_pr3.*
 - 12: gb_ro.*
 - 13: gb_sts.*
 - 14: gb_sy.*
 - 15: gb_un.*
 - 16: em_fun.*
 - 17: em_hum1.*
 - 18: em_hum2.*
 - 19: em_in.*
 - 20: em_om.*
 - 21: em_or.*
 - 22: em_ov.*
 - 23: em_pat.*
 - 24: em_ph.*
 - 25: em_pl.*
 - 26: em_ro.*
 - 27: em_sts.*
 - 28: em_sy.*
 - 29: em_un.*
 - 30: em_v1.*
 - 31: gb_htg1.*
 - 32: gb_htg2.*
 - 33: gb_in1.*
 - 34: gb_in2.*
 - 35: em_bal.*
 - 36: em_ba2.*
 - 37: em_hum3.*
 - 38: em_hum4.*
 - 39: gb_pr4.*
 - 40: gb_htg3.*
 - 41: gb_htg4.*
 - 42: gb_htg5.*
 - 43: gb_htg6.*

- 44: gb_htg7.*
- 45: em_htg1.*
- 46: em_htg2.*
- 47: em_htg3.*
- 48: em_hum5.*
- 49: gb_pl3.*
- 50: gb_pr5.*
- 51: gb_htg8.*
- 52: gb_htg9.*
- 53: gb_htg10.*
- 54: gb_htg11.*
- 55: gb_htg12.*
- 56: gb_htg13.*
- 57: gb_htg14.*
- 58: gb_in3.*
- 59: gb_htg15.*
- 60: gb_htg16.*
- 61: gb_htg17.*
- 62: em_htg4.*
- 63: em_htg5.*
- 64: em_htg6.*
- 65: em_htg7.*
- 66: em_hum6.*
- 67: gb_htg18.*
- 68: gb_htg19.*
- 69: gb_htg20.*
- 70: gb_htg21.*
- 71: gb_htg22.*
- 72: gb_htg23.*
- 73: gb_htg24.*
- 74: gb_htg25.*
- 75: gb_htg26.*
- 76: gb_htg27.*
- 77: gb_htg28.*
- 78: gb_htg29.*
- 79: gb_htg30.*
- 80: gb_htg31.*
- 81: gb_v11.*
- 82: gb_v12.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	484.4	29.0	2043	7	AB013598
2	446.8	26.7	1458	7	AB013597
3	433.2	25.9	1507	7	AB013596
4	289	17.3	1864	7	AB027455
c 5	217.8	13.0	84203	49	AC005106
c 6	217.8	13.0	103223	8	AC007153
c 7	192.6	11.5	198944	8	ATCHRIV38
c 8	192.6	11.5	200576	7	ATFCA0
c 9	166.8	10.0	1589	8	ATG81293
c 10	163.6	9.8	195068	8	ATCHRIV42
c 11	163.6	9.8	205065	7	ATFCA4
c 12	162.2	9.7	197419	8	ATCHRIV41
c 13	131.4	7.9	1594	7	AB000623
14	116	6.9	1440	49	AF190634
15	105.8	6.3	80818	7	AB019232
16	103.6	6.2	568	5	A62510
17	103.6	6.2	568	5	A72383
18	103.6	6.2	1459	5	A62525
19	102.4	6.1	94786	49	AC002333
20	102	6.1	1513	5	A62523
21	102	6.1	1513	5	A72396
22	102	6.1	1969	5	A62529
23	101.8	6.1	1416	7	M2E1AGLU
24	100.4	6.0	566	5	A62509

- AB013598 Verbena h
- AB013597 Perilla f
- AB013596 Perilla f
- AB027455 Petunia x
- AC005106 Genomic s
- AC007153 Arabidops
- AL161538 Arabidops
- 297335 Arabidops
- U81293 Arabidops
- AL161542 Arabidops
- 297339 Arabidops
- AL161541 Arabidops
- AB000623 Nicotiana
- AF190634 Nicotiana
- AB019232 Arabidops
- A62510 Sequence 15
- A72383 Sequence 15
- A62525 Sequence 30
- AC002333 Arabidops
- A62523 Sequence 28
- A72396 Sequence 28
- A62529 Sequence 34
- L34847 Zea mays 1A
- A62509 Sequence 14

25 100.4 5.0 566 5 A72382 A72382 Sequence 14
 c 26 96 5.7 122358 8 AC002396 AC002396 Arabidops
 27 91.2 5.5 99188 49 AC006533 AC006533 Arabidops
 28 91 5.4 48282 53 AC024228 AC024228 Arabidops
 29 90.6 5.4 1431 49 AF165148 AF165148 Petunia x
 c 30 89 5.3 107491 49 AC002391 AC002391 Arabidops
 31 86.4 5.2 107200 49 AC006551 AC006551 Arabidops
 32 86 5.1 1732 7 AB033758 AB033758 Citrus un
 33 83.8 5.0 432 5 A63526 A63526 Sequence 7
 34 82.4 4.9 955 7 AB012115 AB012115 Vigna mun
 35 81.6 4.9 1665 7 AB009370 AB009370 Vigna mun
 36 80 4.8 1588 5 A62526 A62526 Sequence 31
 c 37 79.6 4.8 1687 7 AB016819 AB016819 Arabidops
 38 78.6 4.7 92577 49 AC006282 AC006282 Arabidops
 39 78.4 4.7 1365 8 AF116858 AF116858 Phaseolus
 40 78.4 4.7 1578 5 E12713 E12713 Solanum mel
 41 78.4 4.7 1578 8 SMGT X77369 S.melonigena
 c 42 78.4 4.7 81642 49 AC005851 AC005851 Arabidops
 43 78 4.7 1536 7 AB027454 AB027454 Petunia x
 44 77.2 4.6 114950 49 AC006522 AC006522 Arabidops
 45 77 4.6 1585 7 D85186 D85186 Gentiana tr

ALIGNMENTS

RESULT 1
 AB013598 2043 bp mRNA PLN 20-MAR-1999
 LOCUS Verbena hybrida HGT8 mRNA for UDP-glucose:anthocyanin
 DEFINITION 5-O-glucosyltransferase, complete cds.
 ACCESSION AB013598
 VERSION 1 GI:4115562
 KEYWORDS HGT8: UDP-glucose:anthocyanin 5-O-glucosyltransferase.
 SOURCE Verbena hybrida petal cDNA to mRNA.
 ORGANISM Verbena x hybrida
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Asteridae; Gentianaceae; Lamiales; Verbenaceae; Verbena.
 REFERENCE 1 (sites)
 AUTHORS Yanazaki, M., Gong, Z., Fukuchi-Mizutani, M., Fukui, Y., Tanaka, Y.,
 Kusumi, T. and Saito, K.
 TITLE Molecular cloning and biochemical characterization of a novel
 anthocyanin 5-O-glucosyltransferase by mRNA differential display
 for plant forms regarding anthocyanin
 J. Biol. Chem. 274 (11), 7405-7411 (1999)
 JOURNAL 99167509
 MEDLINE 2 (bases 1 to 2043)
 REFERENCE Yanazaki, M., Saito, K. and Gong, Z.
 AUTHORS Direct Submission
 TITLE Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami
 Yanazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
 Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan
 (E-mail: mamiy@p.chiba-u.ac.jp, Tel.:81-43-290-2905,
 Fax: +81-43-290-2905)

FEATURES

SOURCE
 1..2043
 /organism="Verbena x hybrida"
 /db_xref="taxon:76714"
 /tissue_type="petal"
 26..1411
 /gene="HGT8"
 26..1411
 /gene="HGT8"
 /codon_start=1
 /product="UDP-glucose:anthocyanin 5-O-glucosyltransferase"
 /protein_id="BA336423.1"
 /db_xref="GI:4115563"
 /translation="MSRAVLLATPPAQGHINPALQFAKRLANADIQVTFSTSYAHR
 KSRITAVYSHLFVSDYDGDGQDGNKYMSEKRGKALSDTLANNVDO
 KSRITAVYSHLFVSDYDGDGQDGNKYMSEKRGKALSDTLANNVDO
 AIHLPGLPVLAQRPLSPFLPSTHERFSLMKLETLLEGEEKPVLVNSFDALPD
 ALKALDIYEMIAIGLIPSAFLDGNKDPDRSFGGLDFEKGSDNDXCLLEWLSTNPRSSV

gene

CDS

polya_site
 BASE COUNT 599 a 354 c 508 g 592 t
 ORIGIN
 Query Match 29.0%; Score 484.4; DB 7; Length 2043;
 Best Local Similarity 63.4%; Pred. No. 6.1e-108;
 Matches 920; Conservative 2; Mismatches 453; Indels 75; Gaps 9;
 QY 31 aattataaaaaaattgttaacaaagccattcttactagcaacattccacagacaag 90
 DB 9 AAAAAATAAAAAATAATGAGCAGACTCAGCTGCTCTTGGCCACATTTCCCGACAG 68
 QY 91 gccacataaaccttctctcagttcgccaaaaggctccctcaacacccggtatcgcgcacc 150
 DB 69 GACACATAAATCCCGCCCTTCAATTCGCAAGGCTCTGCGAATGC-----CGACATTC 122
 QY 151 aagtcacattctcacaggtatatacgcattgagacgcatg---cgcttcgaaacgcatc 207
 DB 123 AAGTCACATTTCTACAGCGCTCTACGCATGGCGCCGATGTCACAGAACGGCGGTGGCT 182
 QY 208 cgagcagcagaatcgtattctcgcatktycagatgttctacgatgtggtctaaagaaag 267
 DB 183 CAACGGGCTCATCAATTTTGTGTGCTTTCCGACGGGTATGACGACGGGTACAGCCGG 242
 QY 268 ggcagcgtggcaaaactacatgctcgagatgagaaagcgcggaagcggcgtctaaagg 327
 DB 243 GAGACGATGGGANGAAGTACATGTCGGAGATGAAGACGAGGATATAAAGCTTCGAGCG 302
 QY 328 acactctattaagctcaacgctgctgcgagtgggaagtgaatgttacaatcgctgagct 387
 DB 303 ATACTCTTTCAGCCCAATAATTCGATCAAAAAGACAGCAAA-----ATCACGT 350
 QY 388 ttgtggttactctcatctatttctggtggcagctgaagtggcggtgaagtcgacgtgc 447
 DB 351 TCGTGGTCTACTCCACCTCTTTCATGGGGGCCCAAGGTGGCGGTGAGTTCCTATCCTCC 410
 QY 448 cgagtgcctcttttggattgaacgctgaagtttctcgtatgtgtactatcttttactctta 507
 DB 411 GGAGCGCGCTACTCTGGATTTGAGCCAGCTACGGTGTGGATATATTTTCTTATTTTCA 470
 QY 508 atgggtatgccgatgatacgtatcggtcgcagatcaaatcccaactgccc---aaacttc 564
 DB 471 ACGGCTATAGCCAGCAAAATCGATCGGGTTCGATGCTATTACTTCCCGGAGGACTCC 530
 QY 565 cgcagctctccaagcaagatctccctcttcttcttcttcttcttcttcttcttcttcttct 624
 DB 531 CAGTGTGGCCCGAGCGTGATTTTACGCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 590
 QY 625 gaacctaatgaagaaagttcgacacgctccgacaaagacgaaagcgaaggtcttga 684
 DB 591 GTTCACTGATGAAGGAGAAATGGAACTTTAGAAAGTGAAGAAACCTTAGGCTTTGG 650
 QY 685 taacacgttgcagcatttagaaacccgaaacaaactcaaacgcatcgacaggtatgaactaa 744
 DB 651 TGAACAGCTTTTGATCGGTGGAGCCTGTGCTCAAGAGGCTTTGATTAAGTACGAGATGA 710
 QY 745 tatccatggcccatataatcccatcgcgatattctcagatggcgaacgacccctcatcaa 804
 DB 711 TTGCAATCGGGCGGTGATTCTCTCCGC---ATCTTGGACGCTAAAGATCTCTCGGACA 767
 QY 805 gcaacaaactcagcgtggagacactcttcagaaaaagccgcatgaaactcatcagcagctggc 864
 DB 768 GGTCTTTCGGCGGAGATTGTTGCGAAGAAAGGTCGAATGACGACGATGTCCTCGAATGCT 827
 QY 865 taactcaaaacccgaaatcatcgtgcttctacgtttctggttcgagcctcctgagcctcc 924
 DB 828 TGAGCACGAATCTCGATCTTCGGTGGTTACGTTTACGTTTCTGTTCTGTTCTGTTCTGTT 887


```

Qy 925 cgaaccaccaatggaagaatagcaatagggctttcagacacacaaatcgccagttctct 984
Dy 888 CGAAGTCGCAAAATGGAAGATAGCAAGAGGCGCTGTAGATTGTTGGGAGCCGCTTTTGT 947
Qy 985 ggggtgataagaagaagaagaagggcgagcaacaagaagcaagcagaagaagaagaagc 1044
Dy 948 GGCTGCTAAGAGCTAAGCAAGGAAGAGTA-----TTGATAAGT 988
Qy 1045 tgcctagctcttttgcctgcaggaactgaacactggaacactggaacactggaacactg 1104
Dy 989 TGCATGAGGAGGTTGA-----AACGAGTGGGAAATTTGATCTTGTGTT 1034
Qy 1105 cacaattgagctgttcagcgcataagtcggtggatgcttcctgtagcgcattgcggttga 1164
Dy 1035 CTCAATTGGAAGTCTCTGAGCATCCCTGTTGGGATGTTCTGCACACACTGCGGGTGA 1094
Qy 1165 attcctctatcgagagcctggcttgggtggtgctgcctggtggtgcttcctcaatggtcg 1224
Dy 1095 ATTGCACTCTAGAGATATATCTTTCCGGGTTCCGATGGTGCTTTTCCGCACTGTTCC 1154
Qy 1225 atcaaggactaagcgaagatgcgaagatgctgagagagtggtgtgagagtcagag 1284
Dy 1155 ATCAAGGACGATCGGAGCTGATGGAGGATGCTGAGGACGGGTGTCAGAGTGNAG 1214
Qy 1285 tgnatgaggaagcggtgtgttgcctgtagcgtgagattgaagagtcgctcgcagagttta 1344
Dy 1215 CTAATGAGGAGGAGTACGCTGTTGATGATGAATTTAGGAGATGATTTGAGGAGGTTA 1274
Qy 1345 t-----aaagatcgagagtgtagaagaagcgaactgtagaaggggtttg 1392
Dy 1275 TGGATGGGAGAGAAAGATAGAAACTTAGAGAGTGTGTCGCAAGTGGAGGATTTGG 1334
Qy 1393 ctaaaagactatgtagaagaactgtagatcatcaatgacaaactggaagaatttatta 1452
Dy 1335 CAGAAAAGCTATGGAGGAAGA---TGGATCTTCAGTTAAACAACCTCAAGGCTCTTTCTTG 1391
Qy 1453 ctaggattat 1462
Dy 1392 ATGAGGTGT 1401

RESULT 2
AB013597
LOCUS
DEFINITION
Perilla frutescens PF3R6 mRNA for UDP-glucose:anthocyanin
5-O-glucosyltransferase homologue, complete cds.
ACCESSION
AB013597
VERSION
AB013597.1 GI:4115560
KEYWORDS
PF3R6; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.
SOURCE
Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to
mRNA.
ORGANISM
Perilla frutescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Gentianaceae; Lamiales; Lamiaceae; Perilla.
REFERENCE
1 (sites)
Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,
Kusumi,F. and Saito,K.
Molecular cloning and biochemical characterization of a novel
anthocyanin 5-O-glucosyltransferase by mRNA differential display
for plant forms regarding anthocyanin
J. Biol. Chem. 274 (11), 7405-7411 (1999)
99167509
MEDLINE
2 (bases 1 to 1458)
Yamazaki,M., Saito,K. and Gong,Z.
Direct Submission
Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami
Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan
(E-mail:mamiyep.chiba-u.ac.jp. Tel:+81-43-290-2905,
Fax:+81-43-290-2905)
LOCATION/Qualifiers

```

```

source
1..1458
/organism="Perilla frutescens"
/cultivar="Shikun"
/variety="crispa"
/db_xref="taxon:48386"
/tissue_type="leaf"
29..1360
/genes="PF3R6"
29..1360
/genes="PF3R6"
/codon_start=1
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase
homologue"
/protein_id="BAA36422.1"
/db_xref="GI:4115561"
/translation="MVRPVLLATFPAGHINPALOPAKELLKAGTDVFPFTSVYAWR
RMANATASAAAGNPPGLDFVAFSDYDDGLKPGDGKRYKSEKARGSEALRULLAND
DVFVYSHLFAWAAEVARLHVPTALVKEPATVLCITHFYNGFADSIDAGSEIQ
LDRLEQSLRPLFLPATPERFLMKLETLDEKAKVLNVTADALPEALTA
IDRYELIGIGLIPSAFLDSDPSETSYGGDLFEKSENNVWLNKPKSVVYVF
GSVLFKPAQMEIEIGKLLACGDFPLWIREQNDGDEEEEEEELSCIGELKMGKI
VSMCSOLEVLAHPALGCFVTCGWNSSAVESLSCGIPVVAWPFOWDTTNAKLEDAWG
TGVRYRMNDEGGVDGCEIERCVENVMVDGDKTKLVRENAIKKTKLARQAM"
polyA_site
1458
/note="16 a nucleotides"
BASE COUNT 353 a 348 c 438 g 319 t
ORIGIN
Query Match 26.7%; Score 446.8; DB 7; Length 1458;
Best Local Similarity 63.0%; Pred. No. 9,3e-99;
Matches 896; Conservative 2; Mismatches 434; Indels 90; Gaps 10;
Qy 57 cgcctattctactagcaacattccccagcacaagggccacataaaccctctctcagttc 116
Dy 38 CGCGCGGTGTGTAGCAACAGTTTCGGCGCAAGGCCACATATAATCCGGCCTCAATTC 97
Qy 117 qccaaaaggctctcaacaccggatacgtcgaccacgaagtcacattcttccagagtgatac 176
Dy 98 GCCAAGAGACTCTAAAGCGG-----CACTGAGCTCACGTTTTCACGAGGGTTAT 151
Qy 177 gcattgagacgtcgcttcgaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 224
Dy 152 GCATGGCGCGCATGTCACACACAGCCTCCGCGCTCGCGGAAACCCACCGGCGCTCGAC 211
Qy 225 ttcgtgcatktycagattctacgatgatgctttaaagaagaagcgacgatggcaaaac 284
Dy 212 TTCGTGGCTTCTCCGACGGCTACGACGAGGGCTGAAGCCGCGGCGGAGGAGCGC 271
Qy 285 tacatcgagagatgagaagcgcggaagcgcccttaaggacacactcttattaaagctc 344
Dy 272 TACATGTCGAGATGAAGACCGCGCTCCGAGGCTTTAAGAAACCTCCTTCTCAACAAC 331
Qy 345 aacgatgctcgatgggaagtgaaattacaatcgctgagctgttggtgtgtactctcat 404
Dy 332 GACGA-----CGTCACCTTCGTCGTCTACTCCAC 361
Qy 405 ctatttcgtggcagctgaagtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 464
Dy 362 CTCCTTTTCATGGCGCGGAGGTGGCGCGTTTGTCCACACGTCGCCGCGCGCTCTCTCG 421
Qy 465 attgaacgcgctacggttttcgatgtgtactatttttactcaatgggtatgcccagatg 524
Dy 422 GTCGAGCCCGCACCGCTGTGCATATACACTTCTACTTCAACGGGTACGACAGCAG 481
Qy 525 atcgatcggggctcagatcaaatccaaactgcccactcttcgcgagctctcccaagcagat 584
Dy 482 ATCGACGCGGTTTCCAAATTAATTCAGCTCCCTCGGCTTCCTCCCTGGAGCAGCGCAGT 541
Qy 585 ctccccctttctactctcttcgcagcccgccgagattcccgaaaccccttaataagaagaa 644
Dy 542 CTTCCGAGCTTTCTGCTGCTCGCGCGGAGAGATTCCCGGTTCATGATGATGAGAGAG 601

```

QY 645 ttgcagcgtccgacaaagaccgaagcgaaggtcttgataaacaacgttcgacgcatta 704
 Db 602 CTGGAACATTTTAGCGGTGAAGAGGGGCAAGATATGGTGAACACCTGTATGTCGGTTG 661
 QY 705 gaaacccgaacaaactcaagccatcgacaggtatgaactaatatcatcgcccatattac 764
 Db 662 GAGCCGATGACCTACCGCTATTGATAGTATGAGTTGATCGGATCGGCCGTGTATT 721
 QY 765 ccatacatgatattctcagatggcaacgacccctctatcaagcaacaatactctcagtgga 824
 Db 722 CCCCCTCCG-CTTCTTTGGACGGCGAAGATCCCTC-----CGAACCTCTTACGGGGC 772
 QY 825 gacctcttcgaagaaagccgatgaaact--tacatgactggctaaactcaaaacccgaa 881
 Db 773 GATCTTTTCGAAAAATCGGAGGAGATAACTGCTGGAGTGTGAACCTGAAGCCGAA 832
 QY 882 tcactggtctgttaacgtttctgctggagccctctcaggtctcccgaaaccccaaatggaa 941
 Db 833 TCTTCGGTGGTATGTCTCTTTGGGAGCTTTTGAGGTTTCCAAAGCACAAATGGAA 892
 QY 942 gaaatagcaatagggtcttcagacacaaatcgccagttctctggtgtgataagaagaac 1001
 Db 893 GAGATTGGAAAGGCTATTAGCTCGGAAGGCCCTTTTATGATGATACGAGACAG 952
 QY 1002 gaagagggcgacgaacgaagcgaagcagaagaagaagaagactgctgagctctttgat 1061
 Db 953 AAGAAATGACGACGCGAAGAGAAAGAAAGAAAGAAAGAG-----TTGAGTTGC 1000
 QY 1062 cgtcacggaactgaactcgggacacgaatcgtgacatggtctcacaattggtattctg 1121
 Db 1001 ATTGGGAATTAAGAAAAATGGGAAAAATAGTGTCTGCTGCTCGCAGTTGGAGTTCTG 1060
 QY 1122 acgataagtcggtgggatacttcgtgacgacatgcggttggaattctgctatcgagac 1181
 Db 1061 CGCGACCCCTGGTTGGGATGTTTCGTGACGCATTTGGGTGGAATCTGGCTGTGGAGAGC 1120
 QY 1182 ctggcttgggtgctgcggtggtgctgtcttcctcaatggttcgacagggactcaatgc 1241
 Db 1121 TTGAGTTCCGGGATTCGGTGGTGGCGGTGCCGAGTGTGATGATCAGACGAGGAATCG 1180
 QY 1242 aagatgacgaagatggtggagagtggtgtagagtcagatgaatgaaggaagggcgc 1301
 Db 1181 AAGCTGATTTAGAGATGCTGGGGACAGGGGTGAGAGTGAAGTAAT---GAAGGGGT 1237
 QY 1302 gttgtgtaggcgtgagatgaaggtgctgctcgaggttat-----aaag 1349
 Db 1238 GGGGTTGATGATGAGATAGAAAGGTGTGGAGATGCTGATGATGGGGGTGACAAG 1297
 QY 1350 gtcagagattgagagaagcgaatgatggaaggtttgcttaagaagactatgat 1409
 Db 1298 ACCAACTAGTAGAGAAATGCGATCAATGAAGACATTGGCCAGACACCAACCCATGGGA 1357
 QY 1410 gaagacgtggtatcatcaatgaacaatctgaagaattttatt 1451
 Db 1358 TAGGA---TGGATCTTCACTCAACAATCTCAAGCCCTTTCTT 1396
 RESULT 3
 AB013596 1507 bp mRNA PLN 20-MAR-1999
 LOCUS Perilla frutescens PF3R4 mRNA for UDP-glucose:anthocynin
 DEFINITION 5-O-glucosyltransferase, complete cds.
 ACCESSION AB013596
 VERSION AB013596.1 GI:4115538
 KEYWORDS UDP-glucose:anthocynin 5-O-glucosyltransferase.
 SOURCE Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to mRNA.
 ORGANISM Perilla frutescens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Asteridae; Gentiananae; Lamiales; Lamiales; Gentianaceae; Perilla.
 REFERENCE 1 (sites)
 AUTHORS Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,

Kusumi,T. and Saito,K.
 Molecular cloning and biochemical characterization of a novel
 anthocyanin 5-O-glucosyltransferase by mRNA differential display
 for plant forms regarding anthocyanin
 J. Biol. Chem. 274 (11), 7405-7411 (1999)
 99167509
 2 (bases 1 to 1507)
 Yamazaki,M., Saito,K. and Gong,Z.
 Direct Submission
 Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami
 Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
 Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan
 (E-mail:mamiyep.chiba-u.ac.jp, Tel:+81-43-290-2905,
 Fax:+81-43-290-2905)
 Location/Qualifiers
 1..1507
 /organism="Perilla frutescens"
 /cultivar="Shikun"
 /variety="crispa"
 /db_xref="taxon:48386"
 /tissue_type="leaf"
 17..1399
 /gene="PF3R4"
 17..1399
 /gene="PF3R4"
 /codon_start=1
 /product="UDP-glucose:anthocynin 5-O-glucosyltransferase"
 /protein_id="BAA36421.1"
 /db_xref="GI:4115538"
 /translation="MVRRVLLATFPAOQHINPALQFAKRLKAGTDTVFTSVVNR
 RMNTSAAGNPPGLDFVAFSDGDDGKPGDGRVMSPMKARGSEALRNLLNNH
 DTFVYVYSHLFWAAEVAESQVPSALLNVEPATVLCIYFYFNGYADEIDAGSDEIQ
 LPTEQRQSRLPTFLPETPERFLMKKELETGDEERAKVLDVNTFPADEPDALTA
 IDRYELIGPLIPSAFDGDPFSYGGDLFKSENNCENWMLDTPKSSVYVVSF
 DRYEPKQAMBEIIGKLLACGRFLAMIREQNDGDEEELSCIGELIKMKGVISF
 VRSQLEVLHAPLQFTVHCWNSVESLSCGVVAVVPOWFDQTTNKLIEDAWGTG
 VRYRMNEGGVGDSIERCEVMDGGEKSLVRENAIKWKTILAREAMGDDGSLNKL
 NAFLHOVARA"
 BASE COUNT 366 a 351 c 448 g 342 t
 ORIGIN
 Query Match 25.9%; Score 433.2; DB 7; Length 1507;
 Best Local Similarity 52.5%; Pred. NO. 28-95;
 Matches 889; Conservative 2; Mismatches 435; Indels 96; Gaps 10;
 QY 57 cgcataatctactagcaacattccagcacaagccacataaaccctctctcgagttc 116
 Db 26 CGCCGCGTGTCTGTAGCAACGTTCTCGCCCAAGGCCACATAATCCCGCCCTCCAATTC 85
 QY 117 gccaaaaggtctctcaacacogatacgtcgacaaagtcacattcttctcagtgatatac 176
 Db 86 GCCAAGAGACTCTTAAAAGCCGG-----CACTGAGCTCACATTTTTCACGAGCGTTTAT 139
 QY 177 qcattgagacgcatgcgcttcgaaacccgacgcagcagaa-----tcgat 224
 Db 140 GCATGCCCGCGCATGGCAACACAGCTCCGCGCTGCCGGAACCCACCGGGCGCTCGAC 199
 QY 225 ttctggcathkycagattcttactgatgttgcttaaaagggcagcagtgaggcaaaac 284
 Db 200 TTCTGTGGCGTTCTCCGACGGCTACGACGAGCGGCTGAAGCCCTTGGCGGACGGAAGCGC 259
 QY 285 tacatgctcggagatgagaagcgcggaacgaagccttaaaagcaactcttattaatgctc 344
 Db 260 TACATGTCGAGATGAAGAGCCCGGCTCGAGGCGCTTAAGAAACCTCTCTTCTCAACAC 319
 QY 345 aacgatgctcgcgtggaagtgatgttaacaatcgcggtgagcttctgtggtgactctcat 404
 Db 320 CACGA-----CGTCACGTTGCTGCTCTACTCTCCAC 349
 QY 405 ctatttctgtggcgagctgaagtgccgtggaagtcgacgtgcccagtgccctcttctgg 464
 Db 350 CTCCTTGTGATGGCGGGGAGGTGGCGGTGAGTCCAGGTCCCGAGCGCCCTTCTCTGG 409

Qy 465 attgaaccgggctacggttttcgatgtgtactatttttaactcaatgggtatgccgatgat 524
Db 410 GTGAGCCGCCACCGTGTGTGCATATATTACTTCTACTTCAACGGCTACGAGACGAG 469
Qy 525 atcgatgcgggtccagatcaatacctacacgtcccaatctctccgagctctccaagaagat 584
Db 470 ATGAGCGCGGTTCCGACGAATTCAGTCTCCCTCGGTTCACCCCTGGAGCAGCGCAGT 529
Qy 585 ctcccccttttctactccctctcgagcccgccgagattccgaaccccttaataagaagaag 644
Db 530 CTTCCGACCTTTCTGTGCGGAGACACGGAGATTCGGTTGATGATGAAGAGAGAG 589
Qy 645 ttgcacacgtctgcacaaagacccgaagcgaaggtctttgataaacacgtctgcacgcat 704
Db 590 CTGGAACCTTTAGACGGTGAAGAGAGCGGAAGCTGTGTGATGACAGCTTTGATCGCTG 649
Qy 705 gaaacgcgaacactcaaacccatcgacagggatgaactaatatccatcgcccatatga 764
Db 650 GAGCCCGATGCTACACGGCTATTGATAGGTATGATGATGATGATGATGATGATGAT 709
Qy 765 ccatcgcgatattctcagatggcaacacccctccatcaagcaacaaatcctcaggtgga 824
Db 710 CCTCTCCG---CTTCTTGACGCGGAGATCCCTC-----CGAAACGCTTACGCGGC 760
Qy 825 gacotcttcgaagaaacccgatgaactta-----catgactggtctaaactcaaacccgaa 881
Db 761 GATCTTTTCGAAATTCGAGGAGATAACTCGCTGGAGTGTGTGGACAGGAAGCCGAAA 820
Qy 882 tcatcggtcttcaacttctcggtcgagcctcctcaggtcccgaaaccccaaatggaa 941
Db 821 TCCTCGGTGTGTATGTCGTCTGGGACGCTTTTGAGCTTTTCCAAAGGCACAAATGGAA 880
Qy 942 gaaatagcaatagggttttcagacacccaaatccgagctctctctggttgataagaagaac 1001
Db 881 GAGATTGGGAAGGGGTATTAGCCCTGCGGAAGCGCGTCTTTTATGGATGATACGAGAACAG 940
Qy 1002 gaagagggcgagaaacgaagcaagcagaagaagaagagagagcgtcgtcagctctttgat 1061
Db 941 AAGAATGACGACGCGCAACGAAGAAGA-----GTTGAGTTGC 982
Qy 1062 cgtcagcgaactgaacgactcggggaataatcgtgacatggtctcaccaatggatgtctg 1121
Db 983 ATTGGGGGAATTGAAAAAATGGGAAATAGTTTCGTGGTCTCGCAGTTGCGAGTCTG 1042
Qy 1122 acgcataagtcggtgggagtcctcgagcagcattcggttggaattctgctatcgagagc 1181
Db 1043 CGCACCTCGGTTGGGATGTTTCGTGACGCAATTCGGGTGGAACCTCGGCTGTGGAGGC 1102
Qy 1182 ctggtgtggtgtgcccgtggtgtgcttctcctaagtgttcgatacgaagggaactaatgcg 1241
Db 1103 TTGAGTTGGGGGTTCCGGTGTGGCGTGGCGAGTGTGTGATCAGACGACGAATGCG 1162
Qy 1242 aagatgatcgaagatgtgagggagtggtgtgagagtcagatggaatgaggaagcggc 1301
Db 1163 AAGCTGATTGAGGATCGGTGGGACACGGGTGAGAGTGAGATGAT---GAAGGGGT 1219
Qy 1302 gttgtgataggcgtgagatgaaggggtgctctcggaggttat-----aaag 1349
Db 1220 GGGGTTGATGGATCTGAGATAGAGAGGTGTGTGGAGATGTTGATGAGTGGGGGTGAGAAG 1279
Qy 1350 agtcgagagttgagagaagcgaatgatgtggaaggggttctggtcgaagagctatgat 1409
Db 1280 AGCAACTAGTGAGAGAAATGCCATAAATGGAAGACTTTGGCCAGAGAGCCATGGGA 1339
Qy 1410 gaagaaogtggatcatcatatgaacatctgaagaatttatt 1451
Db 1340 GAGGA---TGGATCTTCACTCAAGAATCTCAACGCCCTTCT 1378

RESULT 4

AB027455 1864 bp mRNA PUN 07-JAN-2000
LOCUS

DEFINITION Petunia x hybrida PH1 mRNA for anthocyanin 5-O-glucosyltransferase, complete cds.
ACCESSION AB027455
VERSION AB027455.1 GI:6683051
KEYWORDS anthocyanin 5-O-glucosyltransferase.
SOURCE Petunia x hybrida (cultivar:Surfinia) corolla cDNA to mRNA.
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE 1 (bases 1 to 1864)
AUTHORS Yamazaki, M. and Saito, K.
TITLE Anthocyanin 5-O-glucosyltransferase
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 1864)
AUTHORS Yamazaki, M. and Saito, K.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences; Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan (E-mail:mamiy@p.chiba-u.ac.jp, Tel:81-43-290-2905, Fax:81-43-290-2905)
FEATURES
Location/Qualifiers
source 1..1864
/organism="Petunia x hybrida"
/cultivar="Surfinia"
/db_xref="taxon:4102"
/tissue_type="corolla"
gene 100..1506
/gene="PH1"
CDS 100..1506
/codon_start=1
/product="anthocyanin 5-O-glucosyltransferase"
/protein_id="BAAB9009.1"
/translation="MVQPHVLTFTPAQGHINPALQFAKNLVKMGLEVTFTSIYAQS RDEKSIINAPKGLNPFQSGDFGDSKDPVFMVSOLRKGSTVKKIILTCSEN GQITCLLYSIFLPMWAEVAREVHI PSALLWSOPATILDIYFNFGHYEKAMANESND PWSIQLPCLPLETRDLPSPYAGKSLRVALPPKELIDLDIAETTPKLTNTF DELEPEALNATEGYKFGYIGFLIPSAFLGNDPLDASFQGLDQNSNDNFWELNSKPN SSVYTSFSGSLMNPISQMEISKGLIDIDPPLMWIKENEKKEENKKGUCIELE KIGKIVPVCQSLVEDLKHPSLVSGHGSNALESACGVPVFAFPQWTDQMTNAQV EDWKSQGVNRVLEKGVSEIEIKRCITELVMDGGEKEELRNKNAKKWELAREVKEG GSSHNKLAFFIDDDVAKGP"

BASE COUNT 603 a 318 c 369 g 574 t
ORIGIN
Query Match 17.3%; Score 289; DB 7; Length 1864;
Best Local Similarity 56.1%; Pred. No. 3.5e-60;
Matches 847; Conservative 2; Mismatches 557; Indels 105; Gaps 12;

Qy 36 aaaaaaaaaatggttaacaaagccatattctactagcaacattccagcacaaagccac 95
Db 88 AAGCAAGAAATATGTCGAGCCTCATGTCATCTTAAACAACATTCACGACAGGCGAT 147
Qy 96 ataaacctctctcagttcgagttcgcaaaaggtctctcaaacccgataogtcgaccaagtc 155
Db 148 ATTAATCCAGCACTTCAATTTGCCAAGATCTTGTCAAGATGGGCATAG-----AAGTG 201
Qy 156 acattctcagcagtgatatacagcagcagcagcgtcttcgaaacccgacgcagc--- 212
Db 202 ACATTTCTACAAGCATTTATGCCCAAGCCGTATGGATGAAATAATCCATCTTAAATGCA 261
Qy 213 ---agcagaatcgtcttggtgcatktycagattcttcagatgattggtcttaagaaagc 269
Db 262 CCAAAAGGATTGAATTTTCATTTCCATTTCCGATGGCTTTGATGAAGGTTTGTGATCATCA 321
Qy 270 gacgatggcaaaactacatgctcgagatgagaaagcgcggaagcgaagcgtcttaagagc 329
Db 322 AAGACCCCTGATTTTACATGTACAACTTCGTAATGTGGAAGTGAACACTGTCAAAAAA 381

QY 330 actcttattaaagctcaacgatgctggagtggaagtgaatgttacaatcgctgagcgttt 389
Db 382 ATAATTCACATTGCTC-----TGAAAATGGACAGCCTATTAACATTGC 423
QY 390 gtgggtactctatctattttcgtggcagtggaagtgcgctggaagtcgacgtgcg 449
Db 424 CTACTTTACTCCATTTTCCCTTCCTTGGGACGAGAGGTAGCAGCTGAAGTTCACATCCCT 483
QY 450 agtgcctctcttttgatgaaacggcgtaacgggtttttcgatgtgtactatattttacttcaat 509
Db 484 TCTGCTCTTCTTTGGAGTCAACGACGCAACAATAATTGGACATATATTACTTCAACTTTCAT 543
QY 510 gggtagccgatgatcgatcggggctcgatca-----aatccaactgcgc 557
Db 544 GGATATGAAAAGCTATGGCTAATGAATCAATGATCCAAATTTGGTCCATCAACTCC 603
QY 558 aatcttcgagctctccaagcaagatctccctcttcttactactccctt----- 606
Db 604 GGGCTTCCACTACTTGGAAATCGAGATCTTCCCTTCATTTTACTTCCCTTATGTTGGTCAAAA 663
QY 607 -cgagcccgcgagatccgaacccctaaagaaagaaagttcgacacgctcgacaaagaa 665
Db 664 GGGAGCTCTCGAGTTGGACTTCCACCATTCAAAGAATTTGATAGACACATTAGATGCTCAA 723
QY 666 cgaagcgaaggtcttgataaacacgttcgacgcatgaaacacgaacgaactcaagcc 725
Db 724 ACCATCTAGATCTTGTGATGATACATTTGATGAATTTAGCAATTTAGCCCTGAGGCATCAATGCA 783
QY 726 atcgacaggtgatgaactaatatccatcgccattatcccatcatcgatattctcagat 785
Db 784 ATTGAAGGTTATAAGTTTATGGAATTTGGACCGTTGATTCCTTCTGCTC---TTCTTGGGT 840
QY 786 gcaacgacccctcatcaagcaacaaatctcgtggagacctcttcagaaagccgat 845
Db 841 GGAATGACCTTTAG-----ATGCTTCATTTGGTGGTGATCTTTTCAAAATTCAAAT 894
QY 846 gaactcatagctggctgaactcaaaacccgaatcatcgctggtttacgtttcgttc 905
Db 895 GA--CTATATGGAATGTTAAATCAAGCCAAATTCATCATGTTGTTATATATCTTTT 951
QY 906 gggagcctcctgaggtcccgaaaccccaatggaagaaatagcaatagggcgttttcagac 965
Db 952 GGGACTCTAATGAATCCATCTATTAGCCAATTTGGAGAGATATCAAAAGGTTGTATGAC 1011
QY 966 accaaatcgacgttctctgggtgataaagaaagaaagagggcgagaaacaaagacaa 1025
Db 1012 ATAGGAAGCGGCTTTTATGGTGATAAAAGAAAATCAAAAAGGCAAGGAAGAGAAT 1071
QY 1026 gcagcgaagaagaagcgtgctgagcttcttgatcgctcaggaactcagcagctcgg 1085
Db 1072 AAA-----AGCTTGGTTGTTATGAAGAAATTTGGAATAATGGAATAATGAGA 1110
QY 1086 aaactcgatcgatggtgtcacatggtatgttctgacgcataaagtcggtgagcttc 1145
Db 1111 AAAATAGTTCCATGCTGTTCACAACTTGAAGTTCTAAAACATCCATCTTTAGGATGTTTT 1170
QY 1146 atgacgcatcgatggaattctgctatcgagcctgctgtggtggtgcctggtg 1205
Db 1171 GTTCTCAITTTGGATGAATTCAGCCTTTAGAGAGTTTACGTTTGGAGTGGCCACTTGTG 1230
QY 1206 tgccttctcaatggttcgatcgaaggaactaatgcaagatgatcgaaagatgtgtggagg 1265
Db 1231 GCATTTCTCTCAATGACAGATCAAAATGACAAATGCCAAACAAAGATGTGTGGAAA 1290
QY 1266 agtgcgtgagagtcagagtcgaagcgcgctgtgttgatagcgcgtgagattaag 1325
Db 1291 AGTGAGTAAAGTGCAGANAATAT---GAGATGGTGTGTTGAAGTCAAGGAAATCAAA 1347
QY 1326 aggtgcgtctcgagggttataaagagtcgagag-----ttgagagaagcgcga 1373
Db 1348 AGGTGATTTGAATTTGGTAAATGGATGGAGGAGAGAAAGGGAAGAAATTTGAGAAAGAAATGCT 1407
QY 1374 atgagtggaaagggtttggctaaagaagcgtatggtgaagaacgtggtatcatcaatgaac 1433

Db 1408 AAGAATGGAAGAATTTGGCTAGAGAAGCTGTGAAGGAAG---GTGGATCTTCACACAAG 1464
QY 1434 aatctgaagaattttattacttaggatttaataagaaatgcctcataagttgtactatat 1493
Db 1465 AATTTAAAGCGCTTTATTATGATGTTGTCCAAGAGGCTTTTAATATTTACAGGCTTTTGCC 1524
QY 1494 atgtattattatt 1504
Db 1525 GTGATATTACT 1535
RESULT 5
AC005106/c
LOCUS
DEFINITION
Genomic sequence for Arabidopsis thaliana BAC T25N20 from
Chromosome 1, complete sequence.
AC005106
VERSION
AC005106.2 GI:6587719
KEYWORDS
HTG.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (bases 1 to 84203)
Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H.,
Conway, A., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M.,
Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and
Ecker, J. R.
Genomic sequence for Arabidopsis thaliana BAC T25N20 from
Chromosome 1
Unpublished
REFERENCE
2 (bases 1 to 84203)
Ecker, J. R.
TITLE
Direct Submission
AUTHORS
Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
JOURNAL
REFERENCE
3 (bases 1 to 84203)
Ecker, J. R.
AUTHORS
Direct Submission
TITLE
Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
JOURNAL
REFERENCE
4 (bases 1 to 84203)
Ecker, J. R.
AUTHORS
Direct Submission
TITLE
Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
JOURNAL
REFERENCE
5 (bases 1 to 84203)
Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C.,
Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chlou, J., Choi, E.,
Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T.,
Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S.,
Mukharasy, N., Nguyen, W., Palm, C., Pham, P., Sakano, H., Schwartz, J.,
Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R.,
Federspiel, N., Theologis, A. and Ecker, J.
Direct Submission
TITLE
Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
JOURNAL
COMMENT
On Dec 16, 1999 this sequence version replaced gi:3366536.
FEATURES
Location/Qualifiers
source
1..84203
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(join(1491..2159,2389..2452,2568..2703,
3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))
/note="5' end of hypothetical protein YUP8H12.1; overlaps

```

T7 end of YUP8H12"
/codon_start=1
/evidence=not_experimental
/product="T25N20.1"
/protein_id="AAC80580.1"
/db_xref="GI:3935137"
/translation="MGKEOKDHSVVEDDKVEAVLHLRKHSPLTLLKQKFCNRRAC
VGLRLIKGDNVKAAKQLRSLCSWRSSIGIESLADDEFAELASGLAYVAGLDDCCR
PVLVRLKQYOKLHTQKQJLTVLVTLEVAISTSRNVQEVPIFLDPAALAAVLLVRL
IVLSFSLNLCSTAHLEFFFEFFIYKINPKLQIISIFHEHTCQJTNQSRRLG
IRTVLSTATMIVMSODFODSDYDDFSKSNYSRVSRLRFTSLKSTDKIGSCASS
RFATVTDGLDKVPCMLTLDTSIKLGHNTGAYISPLNARSFASPAARSEPF3
GPRFSFASTPMFARTDRHSIGTLDPRIPRPFQSPALFRRESHVSKSEPRDS
FVQFLKRYRPPDYRMYRMRPGLAGLVISVSTQIRRHVLSQRF"
join(9304).11268,11387.11513,11607.11692,12103.12341,
12698.12784,13257.13356,13451.14263)
/notes="unknown; similar to ESTs dbj|C26511 and
gb|ai054650"
/codon_start=1
/evidence=not_experimental
/product="T25N20.2"
/protein_id="AAC80581.1"
/db_xref="GI:3935138"
/translation="MEGGGSGRSRIASNTPSRLKESDGVGRTVSSTETRONVQR
SRMVSSESDPMKPPRSQVDRKTGAKKVKRDEHVDHRCYVRRNEASG
SFMKNKLIDFEDEYDFDSANLMKRFDNGSVGRSSFSRRVDSVGRSGGR
EGLDRNRNTPVNCSCASQEDSSSDSDPMRVQGINGVLKVVNKNKTNTLAASI
NPRDAEYFPPSRKAQRRENVVAPPRKNNVDNNSSESDMSKRKKSEYS
KKFEFNTKSTPELVNPDVREERRGGGTQKQLRERIKGLMDAGWTIDYKP
RMSQYLDVAVNPGTAYWSIIKAYDALLKQDKGVDRPRKDTAAVASVSEIYN
KADAKELHRSKSLYNNARPSCGSDSHYLGHRKTKIGRCTLLVRSDKKNPAI
NGFNYSGKPTLLSMLTESGVOLRQVOYMRRCGAKVMLEQWITREGIHDCCKSL
TVSFEIHAGSKSCQPPNIYLSGASLQCOQVRANWOKDATNALAQVTDGDDPN
DDACIGDGDILICDCGFSYTHQNLGMLVLPSPDHMCNCTCFCDAAVASGGKD
GNFISLSCGCHERRHOLCINDEAHKVSFGSASSFGPKCLEFLKQLYLGKTE
IEGYSYLHHRVDTSDTNSQSAQRIENNSKLVAGLAIMDECELPIDRRSGVDLI
GWRCLFPAISNARKYKLEKVIIPAIPDFLHAWTGNFGFTPLDSDVRKMRSLNTLV
PFGIDMLQKLLHDENIIPAAAGDAMIEVETKSEPTSSVEIGPYAVGDEFDVAD
KAANYKDILASDENILVSVETAMGTICKPKOELSRHFERGESISSPCOITLKST
KXVLGHICDVGSCCEGLDVAVEDASALLSQEIQOQASAKFVNNLSLSIRGSS
DLSSLSQVRSEQFSSMLDGVPSCKDYNILVPGAKLDSKODAFADGFL"
complement(join(14820.14957,15038.15420,15643.15721))
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T25N20.3"
/protein_id="AAC80582.1"
/db_xref="GI:3935139"
/translation="MGFLVAVMNFSPTLVHHHMKSPQCQNEKLRQGTSSLFDRRGF
LKCYYGASSMATTEFSGLAQASSEKLDGEGVVGAFKTLDPNERTKSKELPKAY
LKSAREVYKTNRESLKENPNKNAKFRSADAAREKIRYLSNWRGQKTAVAGEESYVEL
ENVIRALAKFYSKAGPSAPLPDVKTEILDOLNKAPEFL"
join(16409.16661,16979.17133,17234.17258,17346.17362)
/notes="unknown; similar to EST emb|Z26799"
/codon_start=1
/evidence=not_experimental
/product="T25N20.4"
/protein_id="AAC80583.1"
/db_xref="GI:3935140"
/translation="MMMDPPPTKEFALPAPVYVWDMKPCVPDDYDARRVGPCKRIL
RKSGVGRPVITAVTAVSLSKVPROILEVSVSTGSLYHEVARPLEEEDSCSETPGPAS
WICSVCRIGPGIAGQGVQDNEITHLSTREHELKROFTNPKGHLAN"
complement(join(17577.18026,18231.18796,18981.19410))
/notes="similar to RNA polymerase II elongation factor
homolog (S59773)"
/codon_start=1
/evidence=not_experimental
/product="T25N20.5"
/protein_id="AAC80584.1"
/db_xref="GI:3935141"
/translation="MFHLVVLIEASIGFVIBCRSYNMDSDIDQSNIGESPHLLRR

```

```

VSPLESGBLPYAPENPNPGDTHMKVGPRIISGKGYFVDRLYLPKYPVLGLDTEILR
KNKFRSLRLQRYIRVHFFPADVQKFTAFSWSIFCRDQGVLPKQKQVLPVYSSE
DMRDDSDTAVCKARNEKCSLMPQCEATLPAFCDCICGKRFKVCDCCLCCKL
ISLEHGYSYIKCAVYSEGHIGCHVAHMCALRAYLAGTIGSGMLGTTEYVCRCD
KKDLFPVHNKFLICQTVYQGVGQILNLGICILRGAQRDNAKELLANCIESTVIKI
FILVYDSDGSAEKNEDTLOSLODTPIGPIFNHEAEMHKEEIEGIVLRALRKAQE
FVYVITAGKHLAOKECILSDLYROLEKESKLSRRVSGTDSANSLMTNLKRLDQIRKEV
TKLEMEVAKGCGRTPRGVLEFFHLNID"
22098..22826
/notes="unknown; similar to ESTs emb|233701 and
emb|Z34709"
/codon_start=1
/evidence=not_experimental
/product="T25N20.6"
/protein_id="AAC80585.1"
/db_xref="GI:3935143"
/translation="MLNTYSLTLELTKTTPRVMMKNFHLCPNSLTKPSSSPSGAT
SDPNRPSILLINNLIDDDSAARHRLKPLIHDFEFSFTSTASTAANSSSSA
SYDDSDNYGFAPDDDDPPDLTAVLASRRFFSPGCSNITSDPLDRDNYDTAIR
LLTGTAVKHYVQSPDPYNDFRKSMQEMIDAVTNAGDLRRYEFHELLLSYLSLNAAD
THKFTIRAFADILVSLSDGHRIS"
complement(join(26300.26503,26613.26880,27004.27259,
27394.27499,27748.27804))
/notes="unknown; similar to ESTs gb|T76156 and gb|N96012"
/codon_start=1
/evidence=not_experimental
/product="T25N20.7"
/protein_id="AAC80586.1"
/db_xref="GI:3935143"
/translation="MAMALNOLFVVLASKPEOKITPESRAIVSCHFPKALWTAGFA
SCVGGDTWQYTKKLPKGLERVALAAGNHLIFVYDHSERLSEKCVFFPFSFTQ
ELLALLLGLIGOVPMQPHLLISLVKMLHECRKSLVRSNRGEAFMOLSKHFY
PESYVGDCRDPQMRWRRTTFTEIASYDDYNATKSQRNPGLPNPSHRRLSGSDA
SKTKTQNSGSDGEMAGEDVLDIVFGSEATESIPAPVLSKASKTQTRKQRAQ
RQKRLREASTNTPQYELA"
complement(join(28447.28740,28976.29135,29206.29943))
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T25N20.8"
/protein_id="AAC80587.1"
/db_xref="GI:3935144"
/translation="MKSRAENQMINRETMMKILVSPSEKSWTSTTTTTTTTTTTS
TEDMRDSNCFDPCRDNACCEICLSNATLDMPLSVQKSLTKLSASNPKS
TVESPTSFDTVTTPASVSRPLKLMISSPKKLTTKSKYFENEQKRTTKERRL
LTVPLKVLIVGLVIGLELGFPSWLVPEVPKPEFTEIVRNAYERTQADLDGVLRLL
LEDFKGFVTSRKFSRCTGSDSKWKINODGPMNLNSKVLVKAIEVSIWGWPLQTS
LFHTGFSSTITVLGSRVTEWTEGKEGYSIHTNTSGTKKSTSVLOLDPWTWLEY
SQSSVDDSSLSLTDIMEHLVFRVAKNVRERYMFFSSRRRLYREASKASATP
T"
join(32077.32407,32715.32956)
/notes="similar to lipid transfer protein (AB007843)"
/codon_start=1
/evidence=not_experimental
/product="T25N20.9"
/protein_id="AAC80588.1"
/db_xref="GI:3935145"
/translation="MNSNPLISAAFLIFSLSSNSTPSTILAQINPCSPSMLSVTGC
TSFLVGKGSFPTSDCCGALKSLGTGMDCLIVTAGVVISIPINTLAISLPACGI
PGVPGCGPASFGPTTSPDSDPDEGTRELFATSSNSKTFCESEFVKRFRGGLL
TSHNSQGLLSVRPLLRLQRLKRLMNHNS"
complement(33000.33155)
/notes="similar to EST emb|Z26808"
/codon_start=1
/evidence=not_experimental
/product="T25N20.10"
/protein_id="AAC80589.1"
/db_xref="GI:3935146"

```

Query Match 13.0%; Score 217.8; DB 49; Length 84203;
 Best Local Similarity 54.2%; Pred. No. 1.1e-42;
 Matches 775; Conservative 2; Mismatches 564; Indels 90; Gaps 13;

QY	35	aaaaaanaaaatggttaacaaacgcacatacttctactagcaacattccagacgaaggcca	94
Db	79169	AGGTAGAAAANAATGGCGCCACCGCAATTTCTACTGGTAACGTTTCCGGCGCAAGGTCA	79110
QY	95	cataaacctctctcagttcgccaaaaggctcctcaacacccgagatacgtcgacaaagt	154
Db	79109	CGTGAACCCATCTCTCCGTTTGTCTGCTGCGCTCATCAAAA--GAACCGCGCACGTGT	79053
QY	155	cacatcttcacgagtgatacgaattgagacgcatcgcttcogaaaacgataccgagcag	214
Db	79052	CACTTTCGTACATTTGTCTCCGCTCTCCCAACATCCATGATCGCAACACCAACAAGT	78993
QY	215	cagaa---tcgattcttgggcaatkycagattcttcagatgatggctttaaagaaagcga	271
Db	78992	CGAAAATCTCTCTTCTTCTTCCTCCGACGGTTTCACGATCGAGGCATTTCCACCTA	78933
QY	272	cgatggcaaaaactacatctcgagatgagaaaagcgcgaacgaagccttaaaagacac	331
Db	78932	CGAAGACGTCAGAAAAGTTCGGTGAATCTCA-----GGTTAACGGCATGAAGCACA	78878
QY	332	tcttattaagctcaacgatgctgcgatgggaagtgaatgttaacatcgcgtgagcttgt	391
Db	78877	TCGGATT---TCATCGAAGCTTACTTAAGAATGGTGACTC-----TCCCGTGACTTGCTT	78828
QY	392	gggtactctcatctatttcgttgggcagctgaaagtgcgcgtgaagtcgcagtcgcgag	451
Db	78827	GATCTACACGATTTCTCAATTTGGGCTCCAAAAGTAGACAGTAGATTTCAACTTCCCTC	78768
QY	452	tgcctctcttggattgaacccggtcaggttttcgatgtgactatttttacttcaatgg	511
Db	78767	CGCTCTCTCTGATGATCAACCGGCTTTGGTTTCAACATCTATTACACTCAITTTATGGG	78708
QY	512	gtatgcgatgatactgatcgggctcagatcgaatacgaactcccaatcttcgcagct	571
Db	78707	AAACNAGTCGGTTTTCAGT-----TACTTAATCTGTTCTTCT	78669
QY	572	ctcgaagcaagatctccctcttctactccctcgagccgcgcgagatccgaac--	629
Db	78668	GGAAATCAGAGATCTTCCATCTTTCTTCCACTTCCACACAAACAAAGGGCGCATACA	78609
QY	630	-ctaatgaaagaaagtctgcacgcctgcacaaagacgaagacgaaggtcttgtataa	688
Db	78608	TGCGTTTCAAGAAATGATGGAGTTTCTCATAAAGAAACCAACCCGAAATTTCTCNTCA	78549
QY	689	caagtctgcacgcattagaaacccgaacactcaagccatcgacaggtatgaactaatc	748
Db	78548	CACTTTCGATTCGCTGGAACACGAGCGCTTAACGGCTTCCCGAATATCGATGGTGGC	78489
QY	749	catcgcccatataatcccatcatcgatatcttcagatggcaacgaacccctctcaagca	808
Db	78488	GGTTGGTCTTTACTTCCACGCGAGATTTCTCTAGGAAGCACCAACAATCA-----	78435
QY	809	caaatctcagctggagacctctcagaaaagccgatgaacttaacatgagcgtgctaaa	868
Db	78436	-----GTTAAAGATCAAGTAGTAGTTATACACTTTGGCTAGA	78399
QY	869	ctcaaaacccgaatcatcggtgttactgttctcggttcggagcctcctgagctcccgaa	928
Db	78398	CTCGAAAACAGAGTCTCTGTATTATTACGTTTCTCTTTTGGAAACAATGGTTGATTGCCAA	78339
QY	929	accccaattgagaataatagcaataggctcttcagacaccaaactcgccagttctctgggt	988
Db	78338	GAAACAGTAGAGAACTAGCGAGAGCACTCATAGAAGGAAACAGACCCTTTTGTGGGT	78279
QY	989	gataa--gaagaaacgaagaggcgacgaaca-aggacaagcgaagaagagaagcct	1045
Db	78278	TATAACTGTATAATCCACAGAGAAACGAAAACAGAGAGGAAGAAGACAGAGATTGA	78219
QY	1046	gctgagctctcttgatcgtcacggaactgaacgactcgggaataatcgatacgtgtgctc	1105
Db	78218	GAAATAGCTGGATTACAGACAGACTTGAAGAGGTTGGGATGATTGTCTCGCTGTGTTT	78159
QY	1106	acaattggatgtctgcagcataaagtcgggtgagctgctcgtgcagcattcgcggttgaa	1165

DB	78158	GCAGATAGAGGTTTAAAGTCACCGAGCCGATAGGTTGTTTGTGACATCATTTGGGTGGAG	78099
QY	1166	tctgcctatcgagagcctggttggctgcccgtggctgctcttctcactggttcga	1225
DB	78098	CTCGACGCTGAGAGTTTGTCTTCCGCTTCCGGTTGTGGCGTTTCCGATGTGTCGGA	78039
QY	1226	tcaaggagactaaatcgaaagatgacgaagatgtgtagagagatggtgtgagagtcagagt	1285
DB	78038	TCAACCGACGACGACCGAAGCTACTGGAAGAAAGTTGGAAGACATGGTGTGAGGGTAAGGA	77979
QY	1286	gaatgagaaagggcgctgttggatagggcgtgagattcaagagtgctcctcgagagttat	1345
DB	77978	GAACAAGGA---TSGTTTGTGAGAGAGGAGAGATCAGGAGGTCTTTTGAAGCCGTGAT	77922
QY	1346	-----aaagagtcgagagtgtagaagaagcgcaatgatgtggaaggttttggctaaaga	1399
DB	77921	GGAGAGAAGTCGGTGGAGTTGAGGAAACGCAAGAAATGGAAGCGTTTAGCGATGGA	77862
QY	1400	agctatggatgaagacgctggatcatcaatgaacatctggaagaatttat	1450
DB	77861	AGC---GGGTAGAGAAGGAGGATCTTCGGATGAACAATGAGGCGCTTTTGT	77814
RESULT	6		
AC007153/c			
LOCUS	AC007153	103223 bp	DNA
DEFINITION	Arabidopsis thaliana chromosome I BAC F320 genomic sequence, complete sequence.		
ACCESSION	AC007153		
VERSION	AC007153.2	GI:4580365	
KEYWORDS	HTG		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.S., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
AUTHORS	Direct Submission		
TITLE	Submitted (24-MAR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
JOURNAL	Submitted (24-MAR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	3 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
AUTHORS	Direct Submission		
TITLE	Submitted (08-APR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
JOURNAL	Submitted (08-APR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	4 (bases 1 to 103223) Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.		
AUTHORS	Direct Submission		
TITLE	Submitted (17-MAY-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
JOURNAL	Submitted (17-MAY-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		

Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 On Apr 9, 1999 this sequence version replaced gi:4508084.
 Bases 1-33654 of clone of F3F20 overlap with bases
 50550-84203 of 'TAMU' clone T25N20, gb|AC005106.
 e-mail for correspondence: arab@sequence.stanford.edu
 Genes with similarity to proteins in the databases are described as
 'putative', '-like' or 'similar to'. Genes that have EST
 similarity but no significant protein similarity are described as
 'unknown proteins'. Genes that are annotated based only on gene
 prediction software are described as 'hypothetical proteins'.
 The software programs used to predict genes include: Grail
 (informatics group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge,
 & A. Salamov, Sanger Centre, <http://genomic.sanger.ac.uk/>), and
 NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).
 Location/Qualifiers

FEATURES

source

```
1. .103223
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F3F20"
1. .33654
/notes="overlap with bases 50550-84203 of 'TAMU' clone
T25N20, gb|AC005106. Features from bases 1 to 29,000 are
annotated on overlapping clone T25N20, gb|AC005106."
complement(29404..40201)
/gene="F3F20.1"
complement(join(29404..29652,29751..29936,30023..30073,
30150..30950,31055..31284,31428..31486,31568..31651,
32740..31858,31981..32100,32252..32411,32506..32695,
32780..32848,32961..33086,33172..33281,33357..33508,
33634..33700,33851..33948,34038..34161,34243..34403,
34491..34575,35347..35441,35543..35680,35806..35892,
35975..36024,36112..36220,36644..36981,37155..37272,
37354..37473,37725..37818,37915..37991,38079..38204,
38319..38445,38591..38682,38787..38879,38964..39080,
39165..39245,39382..39505,39608..39729,39823..39923,
40015..40201)
/gene="F3F20.1"
/notes="Highly similar to putative callose synthase
catalytic subunit; Highly similar to cotton putative
callose synthase catalytic subunit, gb|AAD25952."
```

/codon_start=1

/protein_id="AAD30609.1"

/db_xref="GI:4836907"

```
/translation="MAQRREPPPPPPRRILRTQTVGSLGEAMLDSEVPVSSLVEIAP
ILRVNEASPNRVAICRYAEKARLDPTSSGRVQFKTALLQRLERENETTL
AGRQSDAREMSQFYQHYKRYIQALLNAADKADRLTKAYQTAAVLEVLKAVNQ
EDVEADENLVLDILIKTOIYVPYNILPDPSQNAIMRLPEIQAAVAALRWTRG
LPTWAGKKLDEDILDWQSMFGKDNVYNOREHILLANVHIQRPKPDQPKL
DRALPTLVKILFRYKWCYLGKSSLMPTIQEQVQOKLLYGLVLLWGEAAN
LRPMCELCYIYHLYGLAGSVPMTGERHVPAYGDEDAFLQKYPTIYQISKEA
KISRGGKSHSVRNVDLNEYFWSIRFLGNWRADADFFCQAEURLERSEIKS
NSGRDMGKVAEIRSEVHFIIRFSDRLWSFYILCQAMIVIAWNGSGLSAIFQGDV
FLKLSVFTTALKAQVLDIALSWKARHSLYKLVYKMGVGAAGAAVWVYMAVT
YAYSKNASGESQTKNFGGSHNSPSELTVALIYLSNMLSALFLFPFIRRYLE
RSDYKIMLMWMSQSFYFSSMQIKPLVGPTDKIMRHSIYWSWEEFPHAKNLYG
VIALSPYLTVYMDTQIWAIVATLVGLGNGAFRRGLRITLGLMSRSRSPGAFN
DCLVPODSDPTKKRRFATESRKFDLPSSKOKEARPAQWNNKIISSEFREDIISD
REWELLVPYNSDPLDLIRNPPFLASKPIPIALDMAKDSNGDKRLKRLAVDSYMT
CAVREYAFNRLINLYVVGREGQVINDIFSRDIEHEIKETITELINLSALPDLYGQ
FVRLYLENLENDKQDQIVILLNMLELVRTDIMEEVEPANIYNFDSQFILKRL
GKKIKIRLHLIVAKESAMDVPNLEARRLTFPSNLFMDPPAPKTRNMLSFVL
TPYFDELVSIFLFGQEQNDGVSILFYLQKIPDENTNELEVRKVCNEERAREDL
EEELRWASRGOTLTUTVRGMYYRKALEOAFLDMAKDELLKGYKALELTSEAS
KSGSILACQCALADMKFTFVVCQOYSIHKRSGDQPAKDLRLMCTYPSIVAVIDE
VQQTRESKYKTEEKIYYSALVKAAPQTKPMDSESQVTLDLQILYKILPGPAILGEG
KPNONHAILFTRGBGLQTLDMNQDNWEEAFMRNLLQFLFKHGGVRCPTILGLURE
```

```
HIPTGSVSLAWFMNSQNSFVTIGQRVLASPLKVRHYCHGPDIFDRHLHTRGNST
LREGNTHHEYIQVGKHDVGLNQISMEFAKIANGNEQTLSDRLRLGRHFRDFFML
SCYFTTIGFSTMLTIVTVYVLYLVLSGLEGLSSQRAFRNPKPLKLAASQ
SVQOICFMALPMMEIEGLERGFHNLIEFVLMQLOLASFLLFTFQGTGTHYGRTLF
HGAAYVGTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTG
ILITYSINFMVYVTLFPLFNGSGFMOKIVDQWNTKNIYNRGQISVGPPEKSWFI
WKELEHLRHVSVRGRTITFLAIPFPIFOYGLVYHLSFKGKNSFWYVYASVFI
LELLIVKGLGVGRBRFSTNFQLLPRIKIGLVFLTFVAILITLALPLTLIKDLFTCM
LAPMTGKGNLLIAQACPLIQOGLINSSVYRTLARGYEIVWGLLLTPTVAFNAWFFV
SEFQTRMLFNQAFSRGLQISRLGKRRSRKNKE"
complement(44167..45812)
/gene="F3F20.2"
complement(join(44167..44362,44438..44542,44760..45258,
45331..45446,45524..45593,45680..45812))
/gene="F3F20.2"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AAD30611.1"
/db_xref="GI:4836909"
/translation="MESMGGGEVRRVNLVYFLSRGSHVDPHLLRVHLSRNGVFLR
DYKWLADARGDAMPDAFWSCKRKYKNGVYVQDLDLITPSIDNEYVYKGESEILL
ESPKEYPNVEKKAWTRNGGIDAEKLOKLTSEKIQESVPFCSQRTATSTVVT
BESTNEGFVLKODPKTVSGRDGSGTENGSDNVEGSPSVSTSSSYTKKSY
SSVRASHVLNRLMKCGGLTDNDVILPLNKSRCAGFANEDERRYQYHQHNRKSF
EGASGIRKMKETIEFCFKPVAPSGPMAPICQCGKLFKPKMHSHKLCRGMKNSSA
NNDLMTNNVTVPQRCRNIPGNPLGHQRVLTTLKE"
complement(47118..50318)
/gene="F3F20.3"
complement(join(47118..48467,48761..49149,49186..49302,
49523..49979,50064..50318))
/gene="F3F20.3"
/notes="similar to Na/H antiporter proteins; N-terminal
half of protein is similar to Na/H antiporter proteins"
/codon_start=1
/protein_id="AAD30610.1"
/db_xref="GI:4836908"
/translation="MSSGAGNVNTPNDIIESRFGKIVCYDLSLFEKREKQWGESG
SHVLPFRFMTLEFANLALVYVYVLYVLPYLPFPFVAQILCGLLFSPVGLNTRFII
ATYLVLPNGHPDKIISGCVMSVALACTNPDILARLADLKLRSRDMGTACCAIV
ADTLVLLVYFGFASFSKSGTWNKMPVILITAIFFVLICFIVRGCIRTAFTKTVKA
GGVYLCGLTDACGVHSTIGAFGLSIPDHIINMIEKLDHDFLSGIMLPFLYIIC
GLADIGFMDLQFTDMVYVICSSFLVITVITSLFHPMRDFAIGALMTKNG
TLASVLNAGRTKALDPMWTFMTIALLVSLVSEPLAFAYKPKKLAKHYKHTVQ
MDEKTEPLVACVHLPVNSGITNLQVSNATQSPSLFAIHLVLTGRTTASLLI
MNDCKPKANFSDRVRAESDQIAETFEAMVNDAMTVQITVSPVATWHEIDICVLA
EDKRVCFIILYKHLTPDGRMGNSHAEINQVLSHAPCSVGLVDRGMARVSE
SFESESKEVAMLVGGDDREALSAYARVVGORVILTVRVFPGREALJSSGKVA
AEYEREKQVDDCEIYEFNFKMNSDYKIEKYNQDQDIATIREMNNNSYDLVYV
GRGNSDSPVTLAGLNDSSPELGTIDGTLKSNFTMHASVLIQQLSQRARAVTA
AAATVWAGVAGTGNLLESAGGADKATRAHEPFPMKSNVEDEDEDEHQYGIHR"
51395..53210
/gene="F3F20.4"
join(51395..52184,52258..53210)
/db_xref="GI:4836910"
/translation="MLTSKPHVILIPILFTITLLSPFSLALPINIWPREFLSWPO
HKAIALSPNTTILAPHOVLSAATRVNHLRSNPSPLSYPVKLMKRYTLNVLVT
VTFDSLPHGVDESYSKLSIPGSPSAHLAHSAGMARGLETFSOMIWTCTSDPCLP
VGIYQDSPLPHGRGVLLDTSYGVDDIMRTIKAMSAKLANLWVHHITDSOSPFLV
LPSELSAANGSGPDMVTITDSKIVQYCEHGVNLPKTYEVKNVQIDIVNQVPSFFHGG
VTCANFMWPAKSWERLASEPGTQNLNPSKTYEVKNVQIDIVNQVPSFFHGG
GDEVIPCCWKTQPAJNSLSSGTLQQLLEYINSTPLVIVSNQRTVYVWEDYLDQ
IKADPSLPKEHTILQTNNGPENTKRIVAAAGYVIVSSSEFYLDGCHGFLGNSDI
YDQEGSGGSCAPKFTQSIYNDIADGLNEERKLVLGGEVALMSEQADSTVDSI
RLWRASALASLSWNSRDERGVKRCGEAVDRMLNRYRMVKRIGIAEPIQLWCLKN
PCNCNVHCAADQ"
53685..55199
/gene="F3F20.5"
```

gene

CDS

gene

CDS

gene

CDS

gene

[illegible]


```

/note="strong similarity to major latex protein type 1,
Arabidopsis thaliana; contains EST gb:A1994471.1, T43693,
A109764, H36764, T43905, A1998305.1, R29829, T22625,
R90142, T41926, T22170, W43041, R65053, AA712594,
AA712597"
/codon_start=1
/product="major latex protein like"
/protein_id="CAB78448.1"
/db_xref="GI:7268111"
/translation="MATSGTIVTEVPLKGAERHYKWRNENLFPDAIGHHIGQVTV
HGEVDSGALKIWTLDGKPEMKERREIDDENMAVTFGLEGHVMEQLKVIDTIL
QFIQSPDIDVCKITMTWEKRADDSPEFSNYMKLVKSLAADMDAHLKA"
4845..5031
/gene="AT4g14060"
/number=1
5032..5329
/gene="AT4g14060"
/number=1
5330..5598
/gene="AT4g14060"
/number=2
complement(9562..9654)
/number=1
9562..15425
/gene="AT4g14070"
complement(join(9562..9654,9750..9807,9883..9998,
10097..10176,10276..10435,10594..10731,11065..11145,
11247..11300,11383..11541,11653..11758,11882..11949,
12156..12305,12439..12554,12626..12834,13288..13360,
13958..14072,14674..14859,15291..15425))
/note="strong similarity to AMP-binding protein - Brassica
napus; Contains Putative AMP-binding domain signature
AA255-266; contains EST gb:A1999282.1"
/codon_start=1
/product="AMP-binding protein"
/protein_id="CAB78449.1"
/db_xref="GI:7268112"
/translation="MASTSSLGSTLLSYGSPSRQFPDFGLISGHESVRIPSRFF
REKVPSPFFLESFGDAALRSSEKAWPDIDWSSAEKYGRVALVDVPHDPLK
LTYKLEQILDFAGLVGLVKADEKIALPADNSCHLVSDOGAVNVYGRSSVEE
LQIYHRSVAIVDNEPFRNRTAEFTSKASLRFILLWGEKSLVQCGQIPYIS
YAEIINQGESRAKLSANDNTSRNQFIDSDTAAIMTSGTGPNKGVMLTRNLL
HQTKLSKVPAGAGDKFLSMLPSHAYERASEYFIFTCGVEQMYTSIRYKDDLRY
QNPYVSVPLVYETLYGQIKQISASSAGKFLALTIKYSMAIMEMKRIYEGMCLTK
EQKPPYIVAFVDLWARIALLPWLHMLAKLKIYKKIHSSIGISKAGISGGCLPI
HYDKFEAGTILQNGYGLTETSPVCARTLSCNVLSAGHPMHGTETKIVDPETNNV
LPFGSGIIVKRGPOVMGYVKNPSTTKQVLNESGFNTGDTGWIAPHHSKGRSRHG
GVIVLGRKADIVLSTGENVEPLEIEEAHRSRVIEQIVVIGQDRRRRLGALIIPNKE
EADVDPEFSKETLSLVQELRWKWSCSPQNGPVLIVDDPFTIDNGLMTPMKIRR
DMVYAKKEIDQLYS"
complement(9655..9749)
/number=1
complement(9750..9807)
/number=2
complement(9808..9882)
/number=2
complement(9883..9998)
/number=3
complement(9999..10096)
/number=3
complement(10097..10176)
/number=4
complement(10177..10275)
/number=4
complement(10276..10435)
/number=5
complement(10436..10593)
/number=5
complement(10594..10731)
/number=6
complement(10732..11064)
/number=6
complement(11065..11145)

/number=7
complement(11146..11246)
/number=7
complement(11247..11300)
/number=8
complement(11301..11382)
/number=8
complement(11383..11541)
/number=9
complement(11542..11652)
/number=9
complement(11653..11758)
/number=10
complement(11759..11881)
/number=10
complement(11882..11949)
/number=11
complement(11950..12155)
/number=11
complement(12156..12305)
/number=12
complement(12306..12438)
/number=12
complement(12439..12554)
/number=13
complement(12555..12625)
/number=13
complement(12626..12834)
/number=14
complement(12835..13287)
/number=14
complement(13288..13360)
/number=15
complement(13361..13957)
/number=15
complement(13958..14072)
/number=16
complement(14073..14673)
/number=16
complement(14674..14859)
/number=17
complement(14860..15290)
/number=17
complement(15291..15425)
/number=18
complement(16137..16174)
/number=1
16137..17732
/gene="A74g14080"
complement(join(16137..16174,16252..17604,17687..17732))
/note="the cDNA A26668 is cognate to this gene; gene A6
for anther-specific protein - beta-1, 3-glucanase homolog;
Contains glycosyl hydrolases family 17 signature
AA274-287; Prokaryotic membrane lipoprotein lipid
attachment site AA8-16; contains EST gb:AW004339.1,
Z18400"
/codon_start=1
/product="A6 anther-specific protein"
/protein_id="CAB78450.1"
/db_xref="GI:7268113"
/translation="MSLAFFLTILVFSSSCSATRFQGHRYMQRKTMLDLASKIGI
NYGRGNLPSPYQSFNFKSKAGHVKLYADPESLTLLSOTNLYVTIVPNHQITP
LSSNOTIADEWVNTNLPYQOTRFLVGNELLSVNSGVSVNLVPMARKLVNSLR
LHGHNKIKVGTPLANDLSLSSPPSNGTFREITGPMPLKELKNTSNYFELNVHP
YFRWSRNPNTSLDFALFQGHSTYDPTQGLVYRNLLDQMLDSVLFPMTKLGFPHML
A1SETGWPNGDIDETGANILNAATYRNKLKMSASPPIGTPSRPGLPITPFFVSLP
NENQKSGSQTRHGLHPDGSPIYDVDTGTQPLTGFNPLPKPTNNVPYKQWCVYP
VEGANETELETETLRMACQSNNTCAALAGRECEPVSIYWHASYALNSYWAQFRNOS
IOCFNGLAHETTNPNDRCRFPSSVTL"
complement(16175..16251)
/number=1
complement(16252..17604)
/number=2

```

Intron complement(17605..17686)
 exon /number=2 (17687..17732)
 exon /number=3 (17733..17788)
 gene /gene="AT4g14090"
 CDS /number=1 (17789..17844)
 CDS /gene="AT4g14090"
 CDS /note="similarity to proton-coupled peptide transporter PEPT, Rattus norvegicus; Contains prokaryotic membrane lipoprotein lipid attachment site AA535-545; Prokaryotic membrane lipoprotein lipid attachment site AA955-965; contains EST gb:AI998193.1"

Query Match 11.5%; Score 192.6; DB 8; Length 19894;
 Best Local Similarity 53.2%; Pred. No. 1.6e-36;
 Matches 750; Conservative 1; Mismatches 550; Indels 108; Gaps 12;

QY 60 catattctactagaacattccagcacaagcccaataaaccctctctctcgagttcgcc 119
 Db 21208 CATATTCTGTTGAACATATCCACGCAAGGTCAATCAACCCGGCGCTTCAACTAGCC 21149

QY 120 aaaaagctctcaacacccgagatacgtgcacaaagtcacattctccagagtgatacgca 179
 Db 21148 AACCGGCTCATCCACCACCGTGCACAC-----GTCAATATCTCCACCGCAGTCTCTGCT 21095

QY 180 ttgagacgcatcgcttggaacagatccgagcagaagaatcgattcttggtgcatktyca 239
 Db 21094 CACCGGATGATGGG---CGAGCGACCTTCCACAAAGGTCTATCTCTGGTGGTTCACC 21038

QY 240 gattctacagatgaggtctaaagaagcgacgatgcaaaactacatgctcgagatg 299
 Db 21037 GATGGATTCCAGCAGCGTCTCAAGTCATCGAAGACCCAGAAATCTACATGTCCGAATC 20978

QY 300 agaaagcggaacgaagccttaaggaacactcttattaagctcaagcagatcgctgag 359
 Db 20977 AAMCGATGTGTTCAACCGCTTGAGACATCATCAAGCCATCTTGACGCCACACC 20918

QY 360 ggaagtgaattgaacatcgctgagcttgggtgactctctctctctctctctctctct 419
 Db 20917 GAAACAGACCTATCACCGGGGTAA-----TCTACTCTGTTCTCTGCTCCGCTGGGTT 20867

QY 420 gctgaagtgccgctgaagtgacgagtcgagtgccctctcttggattgaacggctacg 479
 Db 20866 TCTACGGTAGCGCGTGAAGTTCACCTCCCACTACACTTCTGGAATGACCCAGTACT 20807

QY 480 gttttcga tgtgatacttttaactcaatggg gatccgagatgatacgatcgaggctca 539
 Db 20806 GTACTAGACATCTACTACTACTTCAACACCTCTTACAAACATCTC---TTGACGTT 20750

QY 540 gatcaaatccaactgcgaatctcccgagctctcccaagcaagatctccctcttcccta 599
 Db 20749 GACCGGATTAATTTACCAAACTGCACTGATACACCGGTGACCTTCCCGTGTTCCT 20690

QY 600 ctccctctgagccc---cgcgagatctccgaacccctaataaagaaagttgcacagctc 656
 Db 20689 CAACCTTCGAAGGCATATACCGTCTGAGTCTTGTGACTTAAGAGAACATATCGAAGTCTC 20630

QY 657 gacaaagacccgaagcagaggtcttgataaacacgcttcgacattacaaacccaacaa 716
 Db 20629 GAAACGGATCAAAACCCCTAGATCTTCTTAAACACATCTCTGCTTTGGGACACGATGT 20570

QY 717 ctcaagccatcgacaggttgaaactaatatccatcgcccatcattatcccatcatcgata 776
 Db 20569 TTAACCTCTGTTGAGAAATCTAAGATGATCCCAATCGGACCGTGTGTTCTTCTC----- 20515

QY 777 ttctcagatggcaagcagccctcatcaagcaacaaatccatcaggtgagacctcttcaga 836

Db 20516 --CTCCGAGGGTAAAC-----GATCTTTTCAA 20489
 QY 837 aaagcagatgaacattacatgactggtgctaaactcaaaacccgaatcatcgctgtttac 896
 Db 20488 TCTTCCACGAGGATTACACGAAATGGTTAGACTCGAAGCTCGAGACATCAGTATTAC 20429

QY 897 gtttcgttcggagcgtctctcaggg---ctcccgaaaccccaaatggaagaaatagaaca 953
 Db 20428 ATTCTTTAGGCACACACCCGATGATTTACAGAAACACATGGAAGCGCTTACTCAC 20369

QY 954 gggctttcagacacacacaaatcgccagttctctgggtgataagaaagaaacgaagggcgac 1013
 Db 20368 GCGTGTGTAGCTACAAACAGACCGTTTTTATGGATCGTGAG----- 20326

QY 1014 gaacaagacgaagcagaagaagaagaagctgctgagctctcttctgactcgcgaact 1073
 Db 20327 -----GGAGAAATAATCCAGAGAGAAAGAAAGAAATCGGTTCTTGAATTTGATCAGAGGA 20273

QY 1074 gaacgactcgggaacaaatcgtagcatggtgctcacaaattggatgttctgacgcataagtcg 1133
 Db 20272 AGTGATCGAGGATTTGGTGTGGGATGGTGTCTCAGACAGCTGTTTGGCGCATTTGCT 20213

QY 1134 gtggagatgcttcgtgacgcattcggttggaattctctgactcagagacgtcgttggtt 1193
 Db 20212 GTGGGATGTTTGTGACTCATTTGTTGGTAAATTCGACGTTGGAGAGTTTAGAGAGTGT 20153

QY 1194 gtcccggtggtgctttctcctcaatggttcgataagggagactaatcgaaagatgacgaa 1253
 Db 20152 GTTCCGGTGTGGGTTTCCGCGAGTTTCTGATCAGTACACGCGGAGCTTGGAG 20093

QY 1254 gatgtgagagagtggtgtgagagtcagagtgatgaatgaggaagcggtcgttggtagag 1313
 Db 20092 GATACGTGAGGATTTGGAGTGAAGGTGAAGGT---TGGGGAGGAAGGAGATGTGGATGG 20036

QY 1314 cgtgagatlaagaggtcgctcgtcggaggttataaagagtcga-----gagttg 1361
 Db 20035 GAGGAGATTAGAAGGTGTTTGGAGAAGGTGATGATGTGTGGAAGAGCGGAGGAGATG 19976

QY 1362 agaaagcgcgaatcgtggaaggttggcttaaacacacacacacacacacacacacacac 1421
 Db 19975 AGAGAATGACAGAAAGTGGAGGCGGATGCTGTTGATGC---GGCAGCGGAAGGTGGA 19919

QY 1422 tcataatgaacaatctgaagaattttat 1450
 Db 19918 CCGTCGGAATTTGAATCTTAAAGGTTTGT 19890

RESULT 8
 ATCA0/C 200576 bp DNA PLN 28-JUN-1999
 LOCUS Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
 DEFINITION No. 0.
 ACCESSION 297335
 VERSION 297335.2 GI:5280985
 KEYWORDS thale cress, thaliana
 SOURCE Arabidopsis thaliana
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 200576)
 AUTHORS Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N., Kreis, M., Kavanagh, T., Entian, K.D., Rieger, M., James, R., Puigdomenech, P., Hatzopoulos, P., Obermaier, B., Duesterhoft, A., Jones, J., Palme, K., Ansoorge, W., Deisen, M., Bancroft, I., Mewes, H.W., Schaeffer, C. and Chalhatzis, N.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 200576)
 AUTHORS EU Arabidopsis sequencing project.
 REFERENCE Direct Submission
 TITLE Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@ips.biochem.mpg.de, mayere@ips.biochem.mpg.de
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@ncs.ac.uk

On Jun 30, 1999 this sequence version replaced gi:2244747.
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with AtFCAL at the 3' end.

FEATURES

Location/Qualifiers
 1..200576
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 1..63289
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="BAC TAMU20K8"
 1753..4428
 /gene="dl3000w"
 join(1753..3194,3261..4428)
 /gene="dl3000w"
 /note="similarity to Cf-2.2, Solanum pimpinellifolium"
 /codon_start=1
 /product="disease resistance Cf-2 like protein"
 /protein_id="CAB10171.1"
 /db_xref="GI:2244748"
 /db_xref="SPTREMBL:O23253"
 /translation="MTITSLCLIFCLSNILVIAKDLCPDRDALLEEFNERSIP
 SPDSLMALLOPTAKWRNTCCSMGICSDPKTVGVVLDLNSDLNGLRNSLSL
 RLQHLQSLDLSNDLSCTLPDSSGNFYLRVNLGLNLCFGEIPTSRLSLYLDL
 SYNDITGELDSNGLKHLRVLSITCKFTGKIPSSGLNLTLYLTDLSWNYFTGEL
 PSMGNLKSRLVNLHRCNFEFGKIPSTGLSGLNLTLDISKNEFTSEGGDSWSSNRL
 TPLQMLNLSLTWVDSNQKAMLPNSMSLSKLEAFDISGNSFGTIPSSFLM
 PSLKLDLNTDFSGPKIGNISSPNOQLYIGENNINGIPRSLKLVGLSALSLS
 FMDTGYDFISFLQKLSRLSDLSGLNLTSSSHLPSSHMHLLILSSCNISQPKFL
 ENOTSLYLDISANOEGQVPELWRLPTLSFTASDNKFSGETPRACEIGLTVLSNN
 NFSSGTPCFEISNKTLSILHRNNSLSGVIPESLHGLYLRSLDVSNRLSGPKSL
 INCXYOLFVNEENRINTFPFSLKSLPNLQLLVLSRNEHGFIFSGSLFSFKLRF
 FDIENKRSGLVDFYFVGMVMSFVDIINDTPGTFVVGDDSEFHKSYVLTIKGLN
 MELSGGFIYITDVSNGRLEGGDIPESIGLKELIVLNMSNAFTGHPPLPSLSLN
 LSLSDLSQNRISGTPGSLGELTFELARNFYSNMLGPIPGQTQIOSQSSSFAENPG
 LACAPIQKCGGEEDEKEEKDGLSWAAATGYVPGFLCGLAIGHLTYSYKRDWF
 MRIFCSFSSPLKK"
 1753..3194
 /gene="dl3000w"
 /number=1
 3195..3260
 /gene="dl3000w"
 /number=1
 3261..4428
 /gene="dl3000w"
 /number=2
 complement(5905..6149)
 /number=1
 5905..7913
 /gene="dl3005c"
 complement(join(5905..6149,6243..6344,6436..6826,
 7236..7913))
 /note="strong similarity to glycine
 hydroxymethyltransferase (EC 2.1.2.1); contains EST
 gb:1565775"
 /codon_start=1
 /product="hydroxymethyltransferase"
 /protein_id="CAB10172.1"
 /db_xref="GI:2244749"
 /db_xref="SPTREMBL:O23254"
 /translation="MEPVSSNGNTSLVSDVDPETHDLIEKREKQRCIELIASENFTS
 FAVIEALGSLNTKYSIEGIPGNRYIGNERTDEIENLCRSRALEAFHCPAANGVNVQ
 PISGSPANFATLQPHDRIIMGLDLPSSGHLTHGYITSGKKIKSITSIYFESLPYK
 VNFTTGYIDYKLEEKALDFRKLKILCGGSAYPRDNDYARFRAIDKVGALLICDMAH

ISGLVNAQEAANPEFYCDVYVTTTHKSLRGRAGNIIFYRKPKPKKGOPEGAVYDFE
 QKINTAFVPAQGGPHNHQIGALAVALKQANTPGFKYAKQVANAVALGNTYMSKGY
 DVIINGTENHLVLDRLPLGTGKNGKVKELCDLCSITLNKNAVFGDSALAPGVRIGA
 PAMTSGRLGVKDFEQIGFSLRAVYTLTIDTQYTKLLKDFNKGVLNNKDLQDLKADY
 ERFASYPEGPGFLMSEMYKD"
 complement(6150..6242)
 /number=1
 complement(6243..6344)
 /number=2
 complement(6345..6435)
 /number=2
 complement(6436..6826)
 /number=3
 complement(6827..7235)
 /number=3
 complement(7236..7913)
 /number=4
 12823..14568
 /gene="dl3010w"
 join(12823..13533,13822..14568)
 /gene="dl3010w"
 /note="strong similarity to adenosylhomocysteinase (EC
 3.3.1.1)"
 /codon_start=1
 /product="adenosylhomocysteinase"
 /protein_id="CAB10173.1"
 /db_xref="GI:2244750"
 /db_xref="SWISS-PROT:O23255"
 /translation="MALLVEKTSRGYKVKMSQADFGRLELELAEVEMPGLMACRT
 EFGSPPPGARIIGSLHMTIQTAVLIETALGAEVRCWSCNIFSTQDHAANAIAARD
 SAAPFAWKGETQBYWCTERALDWDGPGDPLIVDDGDATILLIEGVKAERIPEKT
 GQVDPSTVDNEFQIVLSILKEGLVDPKYHKMKRELGVSEETTVGKRLYQMQQ
 NGTLLFPAINVDSVTKSFNDFYGRHSPLDGLMRADINIAKVAVICGYGDVKGK
 CAAMATGARVIVTEIDPICALQALMEGLQVLTLEDVVSSEADIFVTITNGDILIMVD
 HMKMKNAIVCNIHGFNEIDMLGETYPGVKRITIKPOTDRWVFPETKAGLIIIVLAE
 GLRMNGCATGCTSPFVMSCTFNQVIAQLELWNEKASKRYEKKYVYLPKHLDEKVAL
 HLGKLCARITKLSKQSDVYSVIPTEGYPKPHYRY"
 12823..13533
 /gene="dl3010w"
 /number=1
 13534..13821
 /gene="dl3010w"
 /number=1
 13822..14568
 /gene="dl3010w"
 /number=2
 complement(16093..16270)
 /number=1
 16093..16501
 /gene="dl3015c"
 complement(join(16093..16270,16317..16501))
 /note="weak similarity to Walleye dermal sarcoma virus gag
 polyprotein"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB10174.1"
 /db_xref="GI:2244751"
 /db_xref="SPTREMBL:O23256"
 /translation="MFNSTALVIFAIFLLISADAFPIPSNGEIDAMLRINSIIGED
 EDLMPTEISRRVYMAQRKNELLITVALDKILITAEAVTLPDALETQTTSTHEGKI
 QETAKIGSFLAEPASKLYD"
 complement(16271..16316)
 /number=1
 complement(16317..16501)
 /number=2
 complement(join(18573..18914,18997..19137,19212..19554,
 19603..19719,19762..20123))
 /note="weak similarity to fusion protein la/1b - sugar
 beet"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB10175.1"
 /db_xref="GI:2244752"

/db_xref="s:PREMBL.O23257",
translation="MDHVSLSPLDEVLVHILSFUTTKAALTSLISKRRNLFTFVPLNL
DIDSVLHPQEKEDREYQKFMKFDVRLAAGNSPIKKLSKLKRTGDFSHRVDG
WISNAGALVETDLILILVYTKLNSRVDMVLAAGDIFLPMKLTVLHSLRVCLVDN
VWNGKGVNYSNLSLITVNYNICLTFSDVTLGFTFVYADYPVVIMONFL
EARLSLVTOOTERARPNIDWDDVVLNVNMEKLIKGRNVQYMYLSPNTLEVL
SCLSSPVENNLKSLTKSHSRGWQVLLRNCPLHETLVLEGILHHVTDKCGVA
CDCVSRKCGSLTSPGVKLEIFGRGTIKEMHIKHPYVFCPLMETNIYIEESGP
PKLRYPQVCFFIAEMEHYKSSSFNVQLLVGGVILYKKTAQ"
complement(18573..18914)
/number=1
gene 18573..20123
/gene="d13020c"
Intron complement(18915..18996)
/number=1
exon complement(18997..19137)
/number=2
Intron complement(19138..19211)
/number=2
exon complement(19212..19554)
/number=3
Intron complement(19555..19602)
/number=3
exon complement(19603..19719)
/number=4
Intron complement(19720..19761)
/number=4
exon complement(19762..20123)
/number=5
misc_feature 20514..20606
/note="EM_ORG:CHSOPSBO Spinach plastid psbB operon with
Query Match 11.5%; Score 192.6; DB 7; Length 200576;
Best Local Similarity 53.2%; Pred. No. 1.6e-36;
Matches 750; Conservative 1; Mismatches 550; Indels 108; Gaps 12;
Qy 60 catattctactagcaacattccacgacagaagccnataaacccctctctcgttgccg 119
Db 81660 CATTACTTGTGTTAACAATCCACGCGCAAGGTCAACAACCGCGGTTCACATAGC 81601
Qy 120 aaaaagctctcaaacgcatagctgcagcaagtcacattcttcacgagtatacgca 179
Db 81600 AACCGGCTCATCCACACGGGTCAACC-----GTCACATACTCCACCGAGTCTCTGCT 81547
Qy 180 ttgagacgcatcgcttgcgaacgcatccgagcagcgaatcgattcttgccatkyca 239
Db 81546 CACCGACGATATGG---CGAGCCACCTTCACAAAAGGTATCTCTTCGCTTGGTTCACC 81490
Qy 240 gatcttcgatgatggcttaaaagaaggcgacgatggcaaaactacatgctcgagatg 299
Db 81489 GATGGATTGACGACGGTCTCAAGTCATTCGAAGACCCAGAAAATCTACATGTCGGAATC 81430
Qy 300 agaaagcgcgaacgaagccttaaaagacaccttattaagctcaacgatgctcgatg 359
Db 81429 AAACGATGTGGTTAAACGCCCTGAGACATCATCAAGCCAATCTTGACGCCACACC 81370
Qy 360 ggaagtgaatgttaaatcgcgtgagcttgggtgactctcatctatttcgtgggca 419
Db 81369 GAANAGACGCTTATCACCGGGGTAA-----TCTACTCTGTTCTCGCTCCGGTGGGTT 81319
Qy 420 gctgaagtgggcgtgaagtcgaagtcgcgagtgccctcttggattgaacgcggtacg 479
Db 81318 TCTACGGTACGCGTGAGTTTCACTCCCAACTACACTTCTGTGGATTGAACGACTACT 81259
Qy 480 gtttcgatgctactatttacttcaatgggtatgcgatgatcatgatcgagggctca 539
Db 81258 GTACTAGACATCTACTACTACTTCAACACCTTTCANACATCTC---TTCGACGTT 81202
Qy 540 gatcaaatccaactgcccaattcttcgagctctccaaagcaagatctccctcttcccta 599
Db 81201 GAACCGATTAAATTACCAGAAATGCCACTGTATACCAACCGGTCACCTCGCTGTTCTT 81142
Qy 600 ctcccttcgaagccc---cgcaagatctccgaacccctaatgaaagaaagttccacacgctc 656

Db	81141	CAACCTTCGAAGGCATTACCGTCAGCTCTTGTGACTCTAAGAGAACATATCGAAGCTCTC	81082
Qy	657	gacaaagacgcaagcgaaggtctctgataaacacgcttcgacgcattagaacccgaacaa	716
Db	81081	GAACCGAATCAACCCCTAAGATCTTCTTAAACACATCTCTGCTTTGGACACGATGCT	81022
Qy	717	ctcaagccaatcgaacaggtatgaactaatatccatcgccccaataatcccatcatcgata	776
Db	81021	TTAACTCTCTGTTGAGAAACTCAAGATGATCCCAATCGACCGTGTGTTCTTC	80967
Qy	777	tctctcgaatggcgaagaccctcatcaagcaacaaatccctacggtggagagcctctcaga	836
Db	80968	--CTCCGAGGGTAAAC--	80941
Qy	837	aaagccgataaacctataatggacttggcttaactcaaaacccgaatactcggctcttac	896
Db	80940	TCCTCCGACGAGGATTAACCAAAATGGTTAGACTCGAAGCTCGAGAGATCACTGATTTAC	80881
Qy	897	gttcgtctcggagagctctcaggg---ctcccccgaacccaaatggaaagaaatagacaata	953
Db	80880	ATTTCCTTAGGCACACGCCCGATGATTACACAGAAACACATGGNACCGCTTACTCAC	80821
Qy	954	gggcttcacgacaccaaatacgcagagttctctcgtggatgaagaagaacgaagagcgagac	1013
Db	80820	GGCGTGTATAGCTACAAACAGACCCGTTTTTATGATCGCTGAG-----	80778
Qy	1014	gaacaagagcaagcagaagaagaagaagctcgtcgagctctcttgaatcgtcacggaact	1073
Db	80779	-----GGAGAAAAATCCAGAGAGAAGAAGAAATCGGTTTCTTGAATTGATCAGAGA	80725
Qy	1074	gaacgactcgggaaaaatcgtgcacatcgtgtcctcaaatctggatctctgaagcaataagtcg	1133
Db	80724	AGTGATCGAGGATTTGGTGGTGGGATGGTGTCTCAGACAGCTGTTTGGCGCATTTGTGCT	80665
Qy	1134	gtggagatctcgtgacgcattcgggttggaattctcctatcgagagcctcggctctggtgt	1193
Db	80664	GTGGGATGTTTGTGACTCATTTGCTGTTTGGAAATTCGACGTTGGAGAGTTTAGAGAGTGT	80605
Qy	1194	gtcccgctgggtgtgcttctcctaattcgtctcgaatacgaaggactaatcgaagaatgatcga	1253
Db	80604	GTTCGGGTGGTGTCCGTTTCCGAGTTGCTGATCAGTGTACAACGGCGAAGCTGTGGAG	80545
Qy	1254	gattgtggagagtggtgtgagagctcagtggaatgaggaagcgcgctgtgtgatagg	1313
Db	80544	GATACGTGGAGGATTGGAGTGAAGTGAAGT---TGGGGAGGAGAGAGATGTGGATGGG	80488
Qy	1314	cgtgagattaaagggtcgctcggagggtataaagagtcga-----gagttg	1361
Db	80487	GAGGAGATTAGAAGTGTTTGGAGAAGGTGATGATGGTGGAAGAAGCGCGGAGGAGATG	80428
Qy	1362	agagaagcgcataatgatgtggaaggttttgctaaagaagctatgatgaagacgttga	1421
Db	80427	AGAGAAATGCAGANACTGGNAGCGATGGCTGTTGATGC---GGACCGGAGAGTGGA	80371
Qy	1422	tcatcaatgaacaatactgaagaatttat	1450
Db	80370	CCGTCGGATTTGAATCTTTAAAGGTTTTGT	80342
RESULT 9			
ATU81293	LOCUS	ATU81293	1589 bp mRNA PLN 02-JUN-1997
DEFINITION	Arabidopsis thaliana UDP-glucose:indole-3-acetate		
ACCESSION	U81293		
VERSION	U81293.1 GI:2149126		

	9	RESULT*
ATU81293		LOCUS
	ATU81293	1589 bp mRNA PLN 02-JUN-1997
DEFINITION	Arabidopsis thaliana up-glucose:indole-3-acetate beta-D-glucosyltransferase (<i>taglu</i>) cDNA, complete cds.	
ACCESSION	U81293	
VERSION	U81293.1	G1:2149126
KEYWORDS	.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Embryophyta; Magnoliopsida; euclerodiales; core eudicotyledons; Rosidae I; Brassicales; Brassicaceae;	
	euclerodiales; Rosidae II; Brassicales; Brassicaceae;	

[illegible]

```

Db 1329 AAGACAGCGGAGGAGTTTAGAGGAATGCCAGGAGGTGGAAGGATTTAGCGGGGAGG 1388
QY 1402 ctatgatgaagaacy 1417
|| || |||| |
Db 1389 CTGTGAGAGAGGAGG 1404

RESULT 10
LOCUS ATCHRIV42/c DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment NO. 42.
ACCESSION AL161542
VERSION AL161542.2 GI:7268303
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 195068)
EU Arabidopsis sequencing project.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
JOURNAL Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG; E-mail:
lemcke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV41 at the 5' end and an
overlap with ATCHRIV43 at the 3' end.
FEATURES
source
location/Qualifiers
1..195068
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
9917..10066
/gene="AT4g15560"
number=1
join(9917..10066,10350..10475,10559..10874,10981..11252,
11340..11613,11691..11998,12081..12362,12441..12725,
12813..12953)
/note="AT4g15560"
Inducible gene required to preserve continuity; light
Contains prokaryotic membrane lipoprotein lipid attachment
site AA452-462; Prokaryotic membrane lipoprotein lipid
attachment site AA591-601; Transketolase signatures
AA102-121; Transketolase signatures AA505-521; contains EST
9b:AW004332.1, AF1997953.1, W43562, N96015, W43388"
/codon_start=1
/product="DEF (CLAL) protein"
/protein_id="CA878598.1"
/db_xref="GI:7268304"
/translation="MASSAAFFPSYIITKGLSTDSCKSTSLSSRSRLVTLDPFCLK
PNNNSNRRAKVCASLAERGEYYSNRPTPLDITINYPDHMKNSVKELKOLSDLR
SDVFNYSKYGGLHGSLSGLVLTVALHYIFNTPQDKILMDVGHOSYPKHILTGRRG
MPTNRTNGSGFTKRGESEHDCFGTGHSTTISAGLMAVGRLGKNNVAVIGD
GAMTAGAYEAMNAGVLDSDMIVILNDNKQVSLPTATLDGSPPPVGSALSRLOS
NPALRELKAVKMTQIGGPMHOLAAKVEYARMSIGTGSLSLFEELGLYYIGPVDG
HIDDLVALILEKYSTTGTGTVLHVITEKGRGYPVAERADPKYHGUVKPDPA"TGROP
KTNTKTSYTYFAEALVAEVEDKDVAIHAAGGGTGLNLPORFFTRCFEDVGIAE
QAVTFPAAGLACGLKFCALYSSFMQRAYDQVHDVLDQKLKLPFRFMDRAGLVGADG
TPHCGAEVTEFMACLPNMIVMAPSDEADLFNMVATAVAIDDRPSCFRPYPRNGIGVAL
PPNGKVPTEIGKRLIKVEGVALDGSVAVQSCIGRAVLEERGLNVTYADARFCK
PLDRALRSIAKSHVEVLITVEEGSIGFGSHVYVFLALDGLDLGKLRWPMVLDPDYI
DHGAPQALAEALMPSHIAATALNLIGAPREALF"
9917..12953
/gene="AT4g15560"
10067..10349
/gene="AT4g15560"
number=1
10350..10475
/gene="AT4g15560"
number=2
10476..10558
/gene="AT4g15560"
number=2
10559..10874
/gene="AT4g15560"
number=3
10875..10980
/gene="AT4g15560"
number=3
10981..11252
/gene="AT4g15560"
number=4
11253..11339
/gene="AT4g15560"
number=4
11340..11613
/gene="AT4g15560"
number=5
11614..11690
/gene="AT4g15560"
number=5
11691..11998
/gene="AT4g15560"
number=6
11999..12080
/gene="AT4g15560"
number=6
12081..12362
/gene="AT4g15560"
number=7
12363..12440
/gene="AT4g15560"
number=7
12441..12725
/gene="AT4g15560"
number=8
12726..12812
/gene="AT4g15560"
number=8
12813..12953
/gene="AT4g15560"
number=9
14413..14436
20338..24557
/gene="AT4g15570"
20338..20373
/gene="AT4g15570"
number=1
join(<20338..20373,20767..20852,20962..21012,21256..21367,
21467..21653..21750,21989..22117,22182..22274,
22625..22672,22957..23015,23122..23206,23286..23765,
24183..24269,24366..24557)
/note="similarity to SEN1 protein, yeast; Contains
ATP/GTP-binding site motif A (P-loop) AA13-20; contains
EST gb:NM5040"
/product="SEN1 like protein"
/protein_id="CA878599.1"
/db_xref="GI:7268305"
/translation="VGLSRKSFVLIQGPCTGKTQITLSILGAIMHATPARVQSGTGD
SEVRKRVLTQEKYNHWGRASPMILGVNPRDAIMPEDGDDGFFFTSGNELKPEVYNA
SRKRLRVLGAVNSALDEIVLRLLSSGLRDENAOQTYTPKIVRIGLKAHISVASVL
DLVAQKRGSVIDPKQGTGTGTDISIRTAILEEAAIYVNGFNFRINVFATISFSQSL
LAKSNRGDVIYIDEAAQAVEPATLIPLATRCQVELVGDGPKQLPATVISTVAQSGY
GTSFMERLQKAGYVPMKLTQYRMHPELIPISOTKPKIVSVLILTPMIASYFTLTIF
LTENQIRSFPSKQFYEGALEGSDIEAQITRDHMKYRCFGPCFFDIHGESQHPGA

```


[illegible]

[illegible]

Qy	575	-----caagcaagatcctccctcttctccctactcccttcgagcccgagatcc	623
Db	567	TTCACTTAGTTCAAGTGAATTTTCCCTTCATTTGTGTTTGATGATGTAAGTCTAATGATG	626
Qy	624	cgaccctaatgaagaagaag-----ttcgacacgctcgacaaagacggaacggaaggt	679
Db	627	GGCAGTTGACTCAATTAAGGCAAAATTCAGATATTGATGAGCAAAATCCAGAAAT	686
Qy	680	cttgataaacacggttcgacgcattagaacccgaacaaactcaagccctcgacagggtatga	739
Db	687	TCCTGTGCAATACTTTTGTGATGCAATTGGAACCTTAATGCTTTGAGGGTTTGGAAATGTAA	746
Qy	740	actcaatatccatcgccccaataatcccatcatcgatatcttcagatggcaacgacccctc	799
Db	747	TATGGTGGAAATGGACCTTATTTCTTCATCTTTCTTGATGA-----GAAAGATC	799
Qy	800	atcaagcaacaactctcagctggagacctctcagaagaacccgatgaactcacatgga	859
Db	800	GAAAAGATAATTTTTCGCGGTGATA-----TGATCGAAGTGAATAATATACATGGA	854
Qy	860	ctggctaaactcaaaacccgaatcatcgcttgattagcttcgctcggaacctcctgag	919
Db	855	GTGGTTGGATGCAAGGCCGAATTAATTCAGTTATTTCATACATCTTGGTAGCTATGCTGA	914
Qy	920	gctcccgaaacccccaaatagaagaatagcaataggctcttcagacacccaaatcgccagt	979
Db	915	GATATCAAGTCAATGGATGGAAGAGATTAGTCAAGGATTGTTGAAATGTGGAGACCATT	974
Qy	980	tctctgggtgataagaagaacacgagagggcgagcaacagagacgaacagagaagaaga	1039
Db	975	TTTGTGGGTGATTAGAAACCCCTAAATGGGGAAAMCCCGAGGAA-----	1020
Qy	1040	gaagctgctgagctcttcttgatcgtcacggaactgaacgactcgggaaaatcgtgacatg	1099
Db	1021	-----AAGTTGACTTGTAAAGATGAATTGGAAAGATAGCGAGAAATAGTGCCTTG	1070
Qy	1100	gtgctcacaaatggatgttctgacgcataaagtctgggtgggagtctcgtgacgcatgcgg	1159
Db	1071	GTGCTCGCAATGGAAAGTCTTAAACACATCGCTCGGTAGGCTGTTCTTGTAATCATTTGTG	1130
Qy	1160	tggaaattcgtcatcgagagcctggctgttggtgtgcgcggtgggtgcttctcctcaatg	1219
Db	1131	ATGGAATTCGACTCTGAGAGCTTAGCTTCTGGTGTGCCAATCGTGGCATGCCCGATTG	1190
Qy	1220	gttcgatcaaggagactaatcgcaagatgatcgaaagtgtgtggaggagtgggtgagaggt	1279
Db	1191	GAATGATCAAAATTTGCAACGCGAAGCTCATTTCAAGATGTTTGGAAAGATTGGTGTGAGAT	1250
Qy	1280	cagagtgaatgaggaagggcgcgctt-----gttgataggcggtgagatgaaggtgcgt	1333
Db	1251	GRACGCTAATAAGAAGGTTATCATTAAGAAGACCGCTTCAGAACTGTATAGAGATCGT	1310
Qy	1334	ctcggagggttataaagtcgagag-----ttgagagaaagcgcaatgatgtggaaggggtt	1390
Db	1311	AATGGGAGATGCTGAGGAAGAGAGAGGAATTGAGAAGAATGCCCAAAATGGGAAGGATTT	1370
Qy	1391	ggctaaagaagctatggatgaagaacgtggatcatcaatgaacaatctgaagaattttat	1450
Db	1371	GGCTAAGGAATCTACGAAGGAANAATAGTTCACTAATGTAATCTTAAAGGCTTATGTTAA	1430
Qy	1451	t	1451
Db	1431	t	1431

RESULT	14
AF190634	
LOCUS	
DEFINITION	Nicotiana tabacum UDP-glucose:salicylic acid glucosyltransferase (SA-gtase) mRNA, complete cds.
ACCESSION	AF190634
VERSION	AF190634.1 GI:7385016
	PLN 02-APR-2000

KEYWORDS	common tobacco.
SOURCE	Nicotiana tabacum
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE	1 (bases 1 to 1440)
AUTHORS	Lee H.-I. and Raskin, I.
TITLE	Purification, Cloning, and expression of a pathogen-inducible UDP-glucose:salicylic acid glucosyltransferase from tobacco
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1440)
AUTHORS	Lee H.-I. and Raskin, I.
TITLE	Direct Submission
JOURNAL	Submitted (28-SEP-1999) Biotech Center, Rutgers University, Foran Hall/Cook College, 59 Dudley RD, New Brunswick, NJ 08901, USA
FEATURES	Location/Qualifiers
source	1..1440
	/organism="Nicotiana tabacum"
	/cultivar="Xanthi-nc"
	/db_xref="taxon:4097"
gene	1..1440
	/gene="SA-GTase"
CDS	34..1413
	/gene="SA-GTase"
	/codon_start=1
	/product="UDP-glucose:salicylic acid glucosyltransferase"
	/protein_id="AAF61647.1"
	/db_xref="GI:7385017"
	/translation="MTQKACHLILPYPAQHINPMLQFSKRLOSGKGVKITIAKTSF LKWOELSTVSEASDGLVDGREGAGTFVAYITREKVGSDTLISOLIGKLTNGCC PVSICVDPLPMAVEVGNNGVAAFTATQSCAVDNLYYHVHGVLKLPPTVDVKEI SVGILLILEADSPVSPVSNPSSRIELMLVNQFSNLENTDVLINSEYELKEVVIDM AKIPIKTIQGTIPSPMYLDKRLPDDREYGLSVFKPMTNACLNLNHPQVSVVYVSFG SLAKLEQMEELALGWSNKNFLMVVRSTEEKLPNNFLEELASEKGLVLSWCPOL QVLEHKSTGFLFCHGNLSNLTALISLGVPMIAMPHSQDPTNAKLVDWDMGTRPKQ DEKGLVRRVIEICIKTVNEEKKKIRENAKKWELARKAVDEGGSDRNIEEFVYSK LVTIASVES"
BASE COUNT	484 a 239 c 318 g 399 t
ORIGIN	
Query Match	6.9%; Score 116; DB 49; Length 1440;
Best Local Similarity	47.9%; Pred. No. 6.6e-18;
Matches 538; Conservative	0; Mismatches 535; Indels 51; Gaps 5;
Qy	353 tgcgatgggaagtgaatgtacaatcgcgctgagctttgttggtactctcatctatttc 412
Db	
Db	315 TGGAAAGTTAACAAATTGTGGTTGTCCTGTGAGTGTGCATAGTTTACATCCATTCTTCC 374
Qy	413 gtgggcagctgaagtggcgcgctgaagtgcagctgcgagtcgacctcttttgattgaacc 472
Db	
Db	375 TTGGGCTGTGTAAGTGGGAAATAAATTTGGAGTAGCTACTGCTGCTTTTTTTCACCAATC 434
Qy	473 ggctacgggttttcgctgctgactatttttacttcaatgggtatgcgcatgatcgaagc 532
Db	
Db	435 TTGTGCAGTGGATACATTTATTACCATGTACATAAGGGGTTCTAAACTTCCCTCCAAC 494
Qy	533 gggtcagat---caaatcactgcgccaatcttcgcgagctctcccaagcagatctccc 589
Db	
Db	495 TGACGTGTGAAGAAATAATCAATTTCTTGGAGTATTAAACAATTTGAGGCATCAGATGTACC 554
Qy	590 ctcttctcactcccttcgagcccgccgagatctccgaacctaatgaagaagaagtctga 649
Db	
Db	555 TAGTTTGTGTTTAATCCTGAATCTTCAAGAATACTTGAATGTTGGTGAATCAGTTCTC 614
Qy	615 cagcgtcgaagaacgcgaagcgaagctcttgataaacacgtctcgcagcattgagaac 709
Db	
Db	615 GAATCTWTGAGAAC-----ACAGATTGGGTCTTAATCAACAGGTTTCTATCAATTGGAGAA 668
Qy	710 cgaaacactcaagccatcgacaggtatgaactaatcattccatcgcccatattatccattc 769

Db 669 AGAAGTAATTGATTGGATGGCCAGATCTATCCAAATCAAGACAAATTTGGACCAACTATACC 728
QY 770 atcgatattcttcagatggcaacgacctccatcaagaacaacaaatctcagcgtggagacct 829
Db 729 ATCAATGTACCTAG-----ACAAGAGCGTACCAAGATGACAAAGAATATGGCCTTAGTGT 782
QY 830 ctccagaaagccgatgaactcatatggactgctgaactcaaaaccccgaaatcatcggt 889
Db 783 CTTCAAGCCCAATGACAAATGATGCCTTAACCTGGTTAAACATCAACCACTTAGTCACT 842
QY 890 cgtttacgtttcgtggagccctcctcagggctcccccgaaccccccaaatggaagaatagc 949
Db 843 AGTATATGTATCATTTGGAAATTTAGCCCAATTTAGAGACAGACCAATGGAAGAAATTAGC 902
QY 950 aatagggcttcagacacccaatcgccagttctctgggtgataagaagaacgaagaggg 1009
Db 903 ATGGGTTTGAGTAAATACCAACAGAACTCTTCTGGGTAGTTAG----- 947
QY 1010 cgacgaacaagagcagaagaagaagaagcgtcgtgagttcttttgatcgctcaggg 1069
Db 948 -----ATCCACTGAAGAATCCAAACTTCCCAACAACCTTTTATAGAGGAATT 992
QY 1070 aactgaagactcgggaaatcgtgacatggtgctcacaattggatggttctgaacataa 1129
Db 993 AGCAAGTGAAGAAGGATAGTCGTGTATCGTGTCCACAATTTACAAGTCTTGGACATAA 1052
QY 1130 gtgtgaggtcttcgtgacgcatgoggttggaattctgctatcgagagcgtgcttg 1189
Db 1053 ATCAATAGGGTGTCTTCACGCGCATGTGGCTGGAATTTCAACTTTGGAAGCAATTAGTTT 1112
QY 1190 tgggtgcccgtgggtgcttctcctcaatggttcgatacgaaggcactaatgccaagatgat 1249
Db 1113 GGGAGTACCAATGATTCGAATGCCACATTTGGTCAGACCCCAACCAANTGCCAGCTTGT 1172
QY 1250 cgaagatgctggaggagtggtgtagagtcagagtgaaatgaggaaggcggcgtt----- 1304
Db 1173 GGAAGATGTTTGGGAGATGGGAATTTAGACCAAAACAAGATGAAGAAAGGATTTAGTAGAAG 1232
QY 1305 -gtgtataggctgagattgaagaggtgctcctcgaggttataaagagtcgagagtgag 1363
Db 1233 AGAAGTTATTGAAGAATCTATTAGATAGTGTGAGGAAAGAAAGAAAGAAAGATTAG 1292
QY 1364 agaaagcgaatgagctgtagaagggtttggtctaaagaagcgtatgagtaagaagcgtggatc 1423
Db 1293 GGAATAATGCAAGAAGAAATGGAAGGAATTTGGCTAGGAAAGCTGTGGATGAAGAGGAAGTTC 1352
QY 1424 atcaatgaacaatctgaagaattttattactaggattattaatg 1467
Db 1353 AGATGAATAATTGAAGAATTTGTTTCCAAAGTTGGTCACTATTG 1396

RESULT 15
AB019232
LOCUS AB019232 80818 bp DNA PLN 20-NOV-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MIL23, complete sequence.
ACCESSION AB019232
VERSION AB019232.1 GI:3869071
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MIL23.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
s.
1 (sites)
REFERENCE
AUTHORS Nakamura,Y.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. I
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 80818)
AUTHORS Nakamura,Y.
TITLE Direct Submission

JOURNAL Submitted (29-OCT-1998) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; 1532-3, Yena, Kisarazu, Chiba 292, Japan
(E-mail:ynakamu@kazusa.or.jp, Tel: +81-438-52-3935,
Fax: +81-438-52-3934)
FEATURES
Location/Qualifiers
1. 80818
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MIL23"
/clone_lib="Mitsui P1"
BASE COUNT 26267 a 14707 c 14502 g 25342 t
ORIGIN
Query Match 6.3%; Score 105.8; DB 7; Length 80818;
Best Local Similarity 55.6%; Pred. No. 2.5e-15;
Matches 243; Conservative 0; Mismatches 167; Indels 27; Gaps 1;
QY 846 gaaacttacatgactgactaaactcaaaacccgaatacctcgttcggttcgttc 905
Db 34402 GATCCTTGATGGAGTGGTTAGACTCGCAGCCAGTTTCTCCGTTGTTTACATCTCATTC 34461
QY 906 gggagcctcctgaggtccccgaaccccaaatggagaataatagcaatagggcttcagac 965
Db 34462 GGGACGGTTGCTTACTTGAACAAAGAACAAATAGACGAGATCGCTTACGGTGTGTTAAAC 34521
QY 966 accaatcccgagttctctgggtgataagaagaacgaaggcgcgaacgaagacaa 1025
Db 34522 GCGACGTTACGTTCTTGTGGGTGATTAGACAAACAGTTAGGTTTCAACAAGAGAAA 34581
QY 1026 gcagaagaagaagaagctgctgagcttcttgcgtcacggaactgacgactcggg 1085
Db 34582 CATG-----TTTTCGGGAAGAAAGTTAAAGGAAGGG 34614
QY 1086 aaaaatcgtgacatggtgctcacaattggtatgtctgcgcataagtcggtgggagtccttc 1145
Db 34615 AAGATCGTTGAATGGTTTCACAGAGAAAGTATTATCTCATCTTCAGTGGCATGTTTC 34674
QY 1146 gtgacgcatcgggttggaattctgctatcgagagcctggccttggtggtgcccgtggcg 1205
Db 34675 GTCACTCACTGTGGATGGAACTCAACGATGGAAGCTGTGCTTCCGGAGTCCCGCGGTT 34734
QY 1206 tgccttcctcaatggctcgatcaagggaactaatgcgaagatgacgaagatgctggaggg 1265
Db 34735 TGTTTCCCTCAATGGGAGATCAAGTCAACGACGCCGCTTTACATGATCGATGTTTGAAG 34794
QY 1266 agtgggtgtagagagtcag 1282
Db 34795 ACGGAGGTGAGGCTAAG 34811

Search completed: August 1, 2000, 13:21:41
Job time: 56763 sec

```

                                GenCore version 4.5
                                Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      August 1, 2000, 14:38:28 ; Search time 3615.02 Seconds
              (without alignments)
              2037.827 Million cell updates/sec

Title:
Perfect score: 1671
Sequence: 1 aacacataaaaaaaaaataa.....ctaaaaaaaaaaaaaaaaaaaa 1671

Scoring table:  IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:      5247842 seqs, 2204914090 residues
                  Total number of hits satisfying chosen parameters:      10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :
1:  em_est1:*
2:  em_est2:*
3:  em_est3:*
4:  em_est4:*
5:  em_est5:*
6:  em_est6:*
7:  em_est7:*
8:  em_est8:*
9:  em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*
```


QY	873	aaaccgaatcatcgtgctgttaccgttcttctgcggagcctcctcgagagctcccgaaaacc	933
Db	83	AAAGCCTAATCATCAATGGTTGTTATCTCATTTGGAGTCTATTGAATTTATCAAGAAC	142
QY	933	caaatggaagaaatagcaatagggctttcagacacaaaatcgccagttctctcgggtgata	992
Db	143	CAAAAGAGGAGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATTTCTTATGGGTAAAT	202
QY	993	agaagaaacgaagaggcgacgaacaaagacagagaaagaaagagagagctgctgagc	1052
Db	203	AGA-----GATCAAGAAGAAGAAAAGAAAGAGAGAAATTAAGTTGC	244
QY	1053	tctttgatcgtcaggaactgaacgaactcgggaaaatcgtgacatggtgctcacaattg	1112
Db	245	ATGATGGA-----ATTAGAGAACGAAGGAAAATAGTACCCTGGTGTTCACAACTT	295
QY	1113	gatgtctgaagcataagtcggtgggatgcttcctgacgcattgcggttggaattctgct	1172
Db	296	GAAGTCCTGACACATCCATTTTAGGATGTTTTGTCTGCCACTGCGATGGATTCGACT	355
QY	1173	atcgagagccctggttgggtgcccgtggtgctcttctcctaagttccatcaagg	1233
Db	356	CTAGAAAGCCATATCGTCAGGTGACCTGTGGTGGCAATTTCTCATTCGACAGATCAAGG	415
QY	1233	actaatgcgaagatgacgaagatggtggaagagtggtgagagtcagagtggaatgag	1293
Db	416	ACTAATGCAAAAGTTGATTGAAGATGTTTGGAGACAGAGTGTCAGGATGAGAGTGAAT---	472
QY	1293	gaagcgcgctgtgtatagcgtgagatgaagagtgctcctcggaggttataaagat	1353
Db	473	GAAGATGTTGGTGTGAGAGTGATGAATATAAANGGTGCATAGAAATTTGTAATGGATGT	532
QY	1353	cgag 1356	
Db	533	GGAG 536	
RESULT	3		
LOCUS	AW442098		
DEFINITION	EST311494 tomato fruit red ripe, TAMU Lycopersicon esculentum clone cLEN21G12 5', mRNA	EST	14-FEB-2000
ACCESSION	AW442098		
VERSION	AW442098.1	GI:6977349	
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Eukaryota: Viridiplantae; Embryophyta: Tracheophyta; Spermatophyta: Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
AUTHORS	Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Lian, Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Roming, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.		
TITLE	Generation of ESTs from tomato fruit tissue		
JOURNAL	Unpublished (1999)		
COMMENT	On Jul 7, 1999 this sequence version replaced gi:5407380. Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU 5 prime sequence.		
FEATURES	Location/Qualifiers		
source	1..649		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="cLEN21G12"		

```

FEATURES
  source
    1. 649
      Location/Qualifiers
        5 prime sequence.
        Email: dfrisch@CLEMSON.EDU
        /organism="Lycopersicon esculentum"
        /cultivar="TA49c"
        /db_xref="taxon:4081"
        /clone="CLEN21G12"

```

/clone_lib="tomato fruit red ripe, TAMU"
 /tissue_type="pericarp"
 /dev_stage="red ripe (7-20 days post-breaker)"
 /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Giovannoni; Fruit were tagged at the
 breaker stage (first sign of lycopene accumulation on the
 blossom end of the fruit) and harvested 7 days
 post-breaker (fully red-ripe), 10 days post breaker, and
 20 days post-breaker (over-ripe). 20 day fruit which
 showed external or internal signs of pathogenesis were
 discarded. Fruit were cut in half and the seeds and
 locules were discarded prior to freezing the pericarp."
 BASE COUNT 217 a 118 c 124 g 190 t
 ORIGIN

Query Match 9.8%; Score 163.6; DB 72; Length 649;
 Best Local Similarity 59.3%; Pred. No. 3.8e-31;
 Matches 390; Conservative 0; Mismatches 229; Indels 39; Gaps 5;
 QY 381 gtgagcttctggttactctctatttctggtggcagctgaagtggcgctgaagtc 440
 Db 1 gtcacgtccctcgctacacactctctgctacgttggcgagctgagtgccgtgaactc 60
 QY 441 gacgtgcccagagtcctctcttctggtgattgaacggcgtacggtttctggtactattt 500
 Db 61 CATATCCCATCCCGGTACTATGATGATCAACGACCAACTGTGCTAGACATATACTACTAT 120
 QY 501 tacttcaatgggtatgcgatgatcgag---tgcgggctcagatca-----a 545
 Db 121 TACTTCAATGGGTATGAGGATGAATGAAGTGTAGTAGTCAATGATCAAAATGGAGT 180
 QY 546 atcccaatgcccaatctccgcagctctccagcaagatctccctcttctcactccct 605
 Db 181 ATCCAAATGCCAAGGCTCCATCTACTAAAGCCCAAGATCTTCATCTTTTGTAGTTCA 240
 QY 606 tcgagccc-----cgagagattccggaacccctaatgaaagaaagtctcgac 650
 Db 241 TCTAGTCAAAAGATGATAAGTATATATGTTTCTCTACCAACATTCAAAGAGCACTAGAC 300
 QY 651 acgtcgcaaaagaccgaagcaggtcttgataaaccagcttcgacgattagaacc 710
 Db 301 ACATTTAGATGGTGAGAAATCCAAAGTACTTGTGAATACATTTGATGCTTAGAGCTA 360
 QY 711 gaacaaatcaagccatgcagcaggtgatgaactaatatccatcgcccttaattcccatca 770
 Db 361 GAGCCCATCAAGCTATTGAAATACAAATTTAATTTGGAATTTGACCATTTGATTCCTTCA 420
 QY 771 tcgatatctcgatggcagcagccctcatcgaagcaacaaatcctacgtgagacctc 830
 Db 421 TC---ATTTCTTGGTGGAAAGAT-----TCATTTGGAATCTCATTTGGTGTGATCTT 471
 QY 831 ttcagaaaagccgatgaacttcacatggactggctgaactcaaaccccggaatcctcggtc 890
 Db 472 TTTCAAAAGTCAAAATGATGACTACATGGAATGGTTTAAACACAAAGCCTAAATCATCAAT 531
 QY 891 gtttacgtttcttcggagcctcctgagctcccgaaaccccaatggagaataagca 950
 Db 532 GTTTATATCTCATTTGGGAGTCTATTGATTTATCAGAAACCCAAAGGAGGAGATTGCA 591
 QY 951 atagggtcttcagaccacaaatcgccaggtctctggtgattgaagaagaaacgagagg 1008
 Db 592 AAAGGGTTGATAGATCCAAAGCCATCTTATGGGTATATAGATCAAGAAGAAG 649

RESULT 4
 AW034683 685 bp mRNA EST 15-SEP-1999
 LOCUS
 DEFINITION
 EST278414 tomato callus, TAMU Lycopersicon esculentum cDNA clone
 cLEC11D5 similar to UDP-glucose:anthocyanin
 5-O-glucosyltransferase, putative, mRNA sequence.
 ACCESSION
 AW034683
 VERSION
 AW034683.1 GI:5893439

KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
 Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1 (bases 1 to 685)
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Viston,T.,
 Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S.,
 Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
 Giovannoni,J.
 TITLE Generation of ESTs from tomato callus tissue
 JOURNAL Unpublished (1999)
 COMMENT On Apr 3, 1998 this sequence version replaced gi:3018868.
 Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@clemson.edu
 5 prime sequence.
 FEATURES
 Location/Qualifiers
 source 1..685
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEC11D5"
 /clone_lib="tomato callus, TAMU"
 /tissue_type="callus"
 /dev_stage="25-40 days old"
 /lab_host="XLI-Blue MRF"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
 of seedlings 7-10 days post-germination were excised, cut
 at both ends and placed on MS medium with no selection.
 Mixed callus was harvested at 25 and 40 days and included
 undifferentiated masses. Tomato Callus EST Library"
 BASE COUNT 219 a 128 c 137 g 201 t
 ORIGIN

Query Match 9.3%; Score 154.6; DB 63; Length 685;
 Best Local Similarity 59.0%; Pred. No. 7.4e-25;
 Matches 378; Conservative 0; Mismatches 224; Indels 39; Gaps 5;
 QY 377 tcgctgagcttctggtgctactctcactatttctggtggcagctgaagtggcggtga 436
 Db 54 TCCTGTGACGCTCCCTCGTCTACACTCTTCTGCTACCTTGGCGAGCTGAGGTAGCGGTGA 113
 QY 437 agtcagcgtgcgagtcgctctcttggattgaacggcgtacggtttctgattgtacta 496
 Db 114 ACTCCATATCCCATCCCGCTTACTATGATTTGATTTCAACAGCAACTGTGTAGCATATACTA 173
 QY 497 ttttactcaatgggtatgcgatgatcgag---tgcgggctcagatca----- 544
 Db 174 CTATTACTTCAATGGCTATGAGTGAATGAAGTGTAGTAGTCAATGATCCAAATTTG 233
 QY 545 ---aatccactgcccactcttctccagctctcccaagcaagatctcccccttttctact 601
 Db 234 GAGTATCCCAATTTGCCAAGGCTTCCATTTACTAAAAAGCCCAAGATCTTCCATCTTTTGTAGT 293
 QY 602 cccttcgagccc-----cgcgagattccgaacccctaatgaaagaaagt 646
 Db 294 TTCATCTAGCTCAAAAGATGATAAGTATAGTTTTCCTCTACCAACATTCAAAGCAACT 353
 QY 647 cgacagcgtcgcaaaagaccgaagcggaaggtcttgataaagcgtctcgacgattaga 706
 Db 354 AGACACATTAGATGGTGAAGAAATCCAAAGGTACTTGTGAATACATTTGATGATGATGA 413
 QY 707 aacgcgaacactcaagccatcgacaggtatgaactaatatccatcgcccatctaaacccc 766
 Db 414 GCTAGAGCCACTCAAGCTATTGAAATACATTTTATTTGGAATTTGACCATTTGATTC 473

```

QY 767 atcaccgatattctcagatggcaacgcgcccctcatcaagaacaacaatctcaggtggaga 826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 474 TTATC- --ATTCTTGGGTGAAAAGAT-----TCATTGGAACTCTTCAATTGGTGGTGA 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 827 cctcttcagaaagccggtgaaacttacatggactggctlaaacctcaaaacccgaatcacc 886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 TCTTTTCAAAAGTCAAAATGATGACTACATGAATGGTTAAACACAAAGCCTAATATCATC 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 887 ggtcgtttacgttttctggaggccctctgaggtcccggaaccccaaatggaagaat 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 AATTCTTATATCTCATTTGGGAGCTCTATTGAATTTATCAAGAAACCAAAAGGAGAGAT 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 947 agcaatagggtcttcagacacaaatcgccagttctctggg 987
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 TGCAAAAGGTTGATAGATCCAAAGGCCATCTTATGGG 685

RESULT 5
AI897621 546 bp mRNA EST 27-JUL-1999
LOCUS EST267064 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION CLD28K21, mRNA sequence.
ACCESSION AI897621
VERSION AI897621.1 GI:5603523
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 546)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Apr 7, 1998 this sequence version replaced gi:3035329.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
Source Location/Qualifiers
1..546
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED28K21"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="Vector: Bluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 196 a 82 c 114 g 154 t
ORIGIN
Source
Query Match 8.2%; Score 137; DB 45; Length 546;
Best Local Similarity 58.7%; Pred. No. 2e-24;
Matches 314; Conservative 0; Mismatches 185; Indels 36; Gaps 3;

QY 632 aatgaagaaagtctcgacgcgtcgcaagaagaccgaagcgaggctcttgataaacac 691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48 ATCAAGAGCAACTAGACACATTAGATGGTGAAGAAAAATCCAAAGGTACTTGTGATAC 107

```

```

QY 692 gttcagcgcattagaaacccgaacaaactcaagccatcgacaggtatgaactaatatccat 751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 ATTTGATGCATTAGAGCTAGAGCCCACTCAAAGCTATTGAAAAATACAAATTTAATGGAAAT 167
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 752 cggccatttaaccatcatcatgatctatctcagatggcaacgacccctcatcaagcaacaa 811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 TGGACCAATGATCTCTTCATC- --ATTCTTGGGTGGAAGA-----TTCATTGGAATC 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 812 atctacggtggagacctcttcagaaaagccgatgaaactatcatggaactggtcaaacctc 871
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 TTCAATTTGGTGGTGTATCTTTTTCAAAAGTCAAAATGATGACTACATGGAATGGTTAAACAC 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 872 aaaaaccgaatcatcgctgctttacgttttcggtcggagcctcctcaggtcccggaacc 931
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 AAAGCTTAATCATCAATTCCTTATCTCATTTGGGAGCTCTATTGAATTTATCAAGAAA 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 ccaaatggaagaataagcaatagggtcttcagacacacaaatcgccagttctctgggtgat 991
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 CCAAAAGAGGAGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCAATCTTATGGGTAAT 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 992 aagaagaacgaagggcgacgaacgaagcaagcagaagaagaagaagaagcgtgctgag 1051
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 AAGAGATCACGAAGAAGAAAAGAGAGAGAAA----- 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1052 cttcttgatcgtcacggaactgaacgactcgggaataatcgtgacatggtcgcacaatt 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 -TTCAGTTTCATGATGGAAATTAGAGAAAGCAAGGGAAAAATAGTACCACTGGTTCACAACT 491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1112 ggaatgtctgacgcataagtcggtggatgcttcgtgacgcattgcggttggaat 1166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 TGAAGTCTGTGACACATCCATCTTTAGGATGCTTTGCTCGCACTGTGGTGAAT 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
AI896013 559 bp mRNA EST 27-JUL-1999
LOCUS EST265456 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLEC13H13, mRNA sequence.
ACCESSION AI896013
VERSION AI896013.1 GI:5601915
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 559)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
Source Location/Qualifiers
1..559
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC13H13"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: bluescript SK(-); Site_1: EcoRI; Site_2:

```

XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 201 a 84 c 112 g 162 t
ORIGIN

Query Match 8.1%; Score 134.6; DB 45; Length 559;
Best Local Similarity 60.0%; Pred. No. 8.3e-24;
Matches 315; Conservative 0; Mismatches 174; Indels 36; Gaps 4;

Qy 632 aatgaagaagaagttcgacacgctcgacaagaacccgaagcgtcttgataaacac 691
Db 71 ATTCAAGAGCAACTAGACACATTAGATGGTGAAGAAATCCAAAGCTACTTCTGAATAC 130
Qy 692 gtccgacgattagaacacgaactcaagccatcgacaggtatgaactatataccat 751
Db 131 ATTGTGATGATTAGACTAGACCCACTCAAGCTATTGAAAATATCAAAATTAATTGGAAT 190
Qy 752 cggccatttaattccatcatgatattctcagatggcaagcaccctctcatcaagcaaca 811
Db 191 TGGACCATGATTCCTTCATC--ATTCTGGTGGAAAGCA-----TTCAATGGATC 241
Qy 812 atctcaggtgagacactcttcagaaaaagccgatgaacttacatgactggctaaactc 871
Db 242 TTCATTTGGTGGTATGCTTTTTCAAAGTCAAAATGATGACTACATGGAATGGTTAAACAC 301
Qy 872 aaacccgaatcatcgctgttcttcgttcgagcctcctcgagctcccgaaacc 931
Db 302 AAAGCCTAAATCAATGTTTATATCTCATTTGGAGTCTATTGAATTTATCAAGAA 361
Qy 932 ccaaatggagaataagcaatagggtcttcagacaccaaatacgccagttctctgggtgat 991
Db 362 CCAAAAGGAGGAGATTGCAAAAGGGTTGATAGACATCCAAAGGCCATCTTATGGGTAAT 421
Qy 992 aagaagaagaagagggcgacgaagcgaagcgaagcgaagcgaagcgtgctgag 1051
Db 422 AAGA-----GATCAGAGAGAGAAAGAGAGAAATTAAGTTG 463
Qy 1052 ctctcttgatcgtccaggaactgaacgactcgggaataatcgacatggctcacaatt 1111
Db 464 CATGATGGA-----ATTAGAGAACAGGGGAAATAGTACATGGTGTTCACAAC 514
Qy 1112 ggatgttcgacgataaagtcggtgggtggtgcttcgtgacgcttg 1156
Db 515 TGAAGCTCTGACACATCCATCTTTAGGATGATTTGCTCGCATG 559

RESULT 7
LOCUS AI899102 478 bp mRNA EST 27-JUL-1999
DEFINITION EST268545 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
CLEP37A16, mRNA sequence.
ACCESSION AI899102
VERSION AI899102.1 GI:5605004
KEYWORDS EST
SOURCE Lycopersicon esculentum
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 478)
AUTHORS Alcala, J., Vrebalov, J., White, R., Matern, A. L., Vision, T.,
Holt, I. E., Liang, F., Upton, J., Ronning, C. M., Craven, M. B.,
Fuji, C. Y., Bowman, C. L., Nierman, W., Fraser, C. M., Venter, J. C.,
Martin, G. B., Tanksley, S. D. and Giovannoni, J. J.
Generation of ESTs from tomato carpel tissue.
Unpublished (1999)
On Feb 18, 1999 this sequence version replaced gi:4296853.
COMMENT Contact: David Frisch
Clemson University Genomics Institute

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.

FEATURES
Location/Qualifiers
source
1..478
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEP37A16"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - tomato Carpel EST Library. Oligo-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5',
and 3' ends located at the EcoRI and XhoI sites,
respectively."

BASE COUNT 172 a 67 c 102 g 137 t
ORIGIN

Query Match 8.0%; Score 134.4; DB 45; Length 478;
Best Local Similarity 60.5%; Pred. No. 9.1e-24;
Matches 310; Conservative 0; Mismatches 166; Indels 36; Gaps 4;

Qy 663 gaaccgaagcgaagctcttgataaacacggttcgacgattagaaacccgaacactcaaa 722
Db 3 GAGAAATCCAAAGTACTTGTGAATACATTTGATGCTTAGAGCTAGAGCCACTCNA 62
Qy 723 gccatgacaggttatgaactaatccatcgcccatcattccatccatcgatattctca 782
Db 63 GCTATTGAAAAATACAAATTAATTGGAATGGACCATGATTCTCTCATC---ATTCTTG 119
Qy 783 gatggcaagcaccctctcatcaagcaacaatacctcagtgagacacctcttcagaaagcc 842
Db 120 GTTGGAAAGAT-----TCATTTGGAATCTTCATTTGGTGTGATCTTTTTCAAAGTCA 173
Qy 843 gatgaacttacatgactggctgaactcaaaaccccggaatcattcggtctgttcgcttccg 902
Db 174 AATGATGACTACATGGAATGGTTAAACACAAAGCCTAAATCATCAATTTGTTATATCTCA 233
Qy 903 ttcggagcctctcgtgaggtcccgaaccccaaatggagaataagcaatagaggtcttca 962
Db 234 TTTGGGAGTCTATTGAATTTATCAAGAAACCAAGAGGAGAGATTGCAGAGGGTTGATA 293
Qy 963 gacacaaatcgccagttctctggtgataagaagaacacgaagagggcgacgacaagaag 1022
Db 294 GAGATCCAAAGGCCATCTTATGGGTAAATAAGA-----GATCAAGAA 335
Qy 1023 caagcagaagaagaagagcgtcgtcgagctctcttcttgatcgtcacggaactggaacgactc 1082
Db 336 GAAGAAAAAGACAGACAGAAATTAAGTTGCATGATGGA-----ATTAGAGAGACAA 386
Qy 1083 gggaaaaatcgtagacatgggtgctcacaaattggatgtctcagcagcataagtcggggaatgc 1142
Db 387 GGGAAAAATAGTACATGGTGGTTGACAACTTGAAGTCTTGACACATCCATCTTTAAGATGT 446
Qy 1143 ttcgtgacgcatgggttggaattctgctat 1174
Db 447 TTTGCTCGCACTGTGGATGGAATTCGACTCT 478

RESULT 8
LOCUS AI489014 500 bp mRNA EST 29-JUN-1999
DEFINITION EST247353 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
CLEP19F7, mRNA sequence.
ACCESSION AI489014
VERSION AI489014.1 GI:4384385

```

RESULT      9
AI487571
LOCUS      520 bp      mRNA
DEFINITION      EST245893 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
                CLED10821, mRNA sequence.
ACCESSION      AI487571
VERSION        AI487571.1  GI:4382942
KEYWORDS       EST.
SOURCE        tomato.
ORGANISM      Lycopersicon esculentum
               Eukaryota; Viridiplantae;
               Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
               Solanaceae; Solanum: Lycopersicon.
REFERENCE     1 (bases 1 to 520)
AUTHORS      Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
               Holt,I.E., Liang,F., Opton,J., Ronning,C.M., Craven,M.B.,
               Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
               Martin,G.B., Tanksley,S.D. and Giovannoni,J.
               Generation of ESTs from tomato carpel tissue
               Unpublished (1999)
TITLE        On Apr 7, 1998 this sequence version replaced gi:3035667.
JOURNAL
COMMENT      Contact: David Frisch
               Clemson University Genomics Institute
               Clemson University
               100 Jordan Hall, Clemson, SC 29634, USA
               Tel: 864 656 4366
               Fax: 864 656 4293
               Email: df@cis.clemson.edu

```

```

FEATURES
source
email: driesche@ECONOM.EDU.
Location/Qualifiers
1. 520
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CJED10D21"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="X11-Ripe MRs'"

```

```

/tissue_type= carpel
/dev_stage=5 days post-anthesis to 5 days post-anthesis"
/lab_host=XLI-Blue MRP"/
/note=vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively.*

```

BASE COUNT 194 a 76 c 107 g 143 t
ORIGIN

Query Match 7.8%; Score 130.4; DB 40; Length 520;
Best Local Similarity 61.9%; Pred. No. 9.5e+23;
Matches 244; Conservative 0; Mismatches 141; Indels 9; Gaps 2;

Qy 632 aatgaagaagattctgacagcgtcgacaagaacgcgaagccgaagtcttgataaacac 691
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 ATTCAAGAGCACATGACACACTAGATGGTGGAAGAATACTCCAAAGGTACTTTGTGAATAC 114
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 692 gtctgcgcattaggaaaccgaactcaagccatcgacgggatgaactaatcccat 751
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 115 ATTTGATGCATTAGAGCTAGAGCCACTCAAAGCTATTGAAAATAFACAATTTAATTGGAAAT 174
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 752 cgcccattaatcccatcatcgatatcttcagatggccaagccccctcatcaagaacaacaa 811
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 175 TGGACCATTGATTCCTTCATC---ATTCTCGGGTGGAAAAAGA-----TTCATTTGGAAATC 225
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 812 atctcacggtggagaccttttcagaagaacgcgatgaacttacatggactggcctaaccctc 871
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 226 TTCAATTTGGTGGTGATCTTTTCAAAGTCAAATGATGACTACATGGAATGGTTTAACAC 285
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 872 aaaccgcgaatcatcggctgctttagcttctcggaagcctcctcaggctccgccgaancc 931
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 286 AAGGCCATAATCATCAATGGTTTATTCATTTTGGGAGTCTATTGAATTTTATCAAGAAA 345
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 932 cc aa at gg aa ga at ag ca at ag gg c t t t c a g a c a c c a a at o g c a g t t t c t c t g g g t g a t 991

```

Db 346 CCAAGAGGAGGATTGCAAAAGGCTTCATAGAGATCCAAAGGCCATCTTATGGGTAT 405
QY 992 aagaagaacgaagggcgacgaacagagcaa 1025
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 406 AAGAGATCAAGAAGAGAGAAAGAAAGAGAGAAA 439
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AI483541
LOCUS EST249390 tomato ovary, TAMU Lycopersicon esculentum cDNA clone 29-JUN-1999
DEFINITION CLED2SM13, mRNA sequence.
ACCESSION AI483541
VERSION AI483541.1 GI:4387465
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 476)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
TITLE Contact: David Frisch
JOURNAL Clemson University Genomics Institute
COMMENT Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.

FEATURES
source
1..476
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TAX96"
/db_xref="taxon:4081"
/clone="CLED2SM13"
/tissue="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5',
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 180 a 64 c 101 g 131 t
ORIGIN

Query Match 7.7%; Score 128.8; DB 40; Length 476;
Best Local Similarity 61.7%; Pred. No. 2.4e-22;
Matches 243; Conservative 0; Mismatches 142; Indels 9; Gaps 2;

QY 532 aatgaagaagaagttgcacacgctgcagaaagccgaagcgaggtcttgataaacac 691
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 55 ATTCAAGAGCAACTAGACACATTAGTGGTGAAGAAATCCAAAGGTACTTGTGAATC 114
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 692 gttcgacgcattagaacccgacaactcaagccatgcagcgggtatgaactaatccat 751
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 115 ATTTGATGCTTAGAGCTAGAGCCACTCAAAGCTATTGAAATAFACAATTAATTTGGAAT 174
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 752 cggccattatcccatcatgattctcagatggcgaagccctcatcaagcaacaa 811
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 175 TGCACATTGATTCCTTCAIC---ATTCTGGGTGGAAAGA-----TTCATTTGGAATC 225
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 812 atocacgdtggagacctcttcagaaagccgatgaaacttacatgactgagcgtaaacctc 871
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 226 TTCAATTGGTGTGATCTTTTCAAAAGATCAAAATGATGACTACATGGAATGGTTAAACAC 285
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

```

```

QY 872 aaaaaccgaatcatcagtgctgtttagcttctcggtggagcctcctcgaggctcccgaaacc 931
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 286 AAAGCCTAAATCATCAATTTGTTATATCTCATTTGGGAGTCTATTGAATATCAAGAAA 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 ccaaatggagaataagcaatagggctttcagacacacacaaatcgccagttctctcgtgggagat 991
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 346 CCAAGAGGAGGAGATTGCAAAAGGCTTGATAGAGATCCAAAGGCCATCTTATGGGTAT 405
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 992 aagaagaacgaagggcgacgaacagagcaa 1025
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 406 TAGAGATCAAGAAGAGAGAAAGAAAGAGAGAAA 439
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AW349414/C
LOCUS AW349414 747 bp mRNA EST 01-FEB-2000
DEFINITION GM210007A20D2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
ACCESSION AW349414
VERSION AW349414.1 GI:6847124
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 747)
AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Kharne,A., Coryell,V.,
Erpelting,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
TITLE On Jul 9, 1999 this sequence version replaced gi:5434916.
JOURNAL Contact: Vodkin, L.O., PI, A Functional Genomics Program for
COMMENT Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888)919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES
Location/Qualifiers
1..747
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone_lib="Gm-r1021"
/tissue_type="root"
/lab_host="XLI0-Gold"
/notes="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; Library Gm-r1021 is a sequence-driven, reracked set
of the original library Gm-cl004 which was prepared from
root cDNA. The mRNA was isolated from entire roots of 8
day old 'Williams' seedlings which were propagated on
paper towels with distilled water. Stragene's cDNA
Synthesis Kit (catalog #200401) was used to synthesize
the cDNA. The Gm-cl004 library was constructed by Dr.
Paul Keim & Virginia H. Coryell, Department of Biology,
Box5640, Northern Arizona University, Flagstaff, AZ
86011, email: paul.keim@na.u.edu, virginia.coryell@na.u.edu.
The contig analysis to select unique genes was performed
by the laboratory of Ernest Retzel, Computational Biology
Centers, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and sequencing by the Keck
Center for Comparative and Functional Genomics,
University of Illinois,

```

http://www.life.uiuc.edu/biotech/keck.html."

```
BASE COUNT      206 a   235 c   97 g   187 t   22 others
ORIGIN

Query Match      7.4%; Score 123.6; DB 71; Length 747;
Best Local Similarity 57.2%; Pred. No. 5.3e-21;
Matches 293; Conservative 0; Mismatches 174; Indels 45; Gaps 2;

QY 851 ttacatgactgctgaactcaaacccgaatcatcggttgcttaccgttcttcggtgag 910
Db 656 TTACCTTGAATGGTGGACTCACAGCTGAGTTATGCTGTTTATGTTTATCTTGGTAC 597
QY 911 cctcctgagctcccgaaaccccaatggaagaatagaactagggcttccagaccaca 970
Db 596 CCTTCTGCTGTGGCTGATACAGATGAAGGACTTGCACGCGCTCTAGATTCGG 537
QY 971 atcgccagttctctggtgataagaagaacgaagggcgacgaacgaagcagacaga 1030
Db 536 ATATCTCTTCTTGTGGGTCATTAGAGATATGCAAGGAATAGAAAGATAACTGCAGAGACGA 477
QY 1031 agaaagaagaagctgctgagctcttcttgatcgtaacgaactgaacgactcggaataa 1090
Db 476 ACTGGAGCAGAGG-----GCTAGAT 456
QY 1091 cgtgacatggtgctcaaatggtatggtctgacgcataagtcgtggtggttcgtgac 1150
Db 455 TGTGAATGTTGTTCTCAGGTGGAGGTCTGTGCAATGTTCTTGGGTGTTTGTAAAC 396
QY 1151 gcatlccggttggaattctgctacgagcgtggtctggtggtgctggtggtgctt 1210
Db 395 GCATTGTGGATGGAATTCGACTATGGAAGCTTTGGGTTCGGGGGTTCCTATGTTGGCGTT 336
QY 1211 tctcaatggttcgataagggataatgcaagatgataagaatgctgtagagagtgag 1270
Db 335 TCCGCAGTGGACAGACACCAAGGAGCAATGCAAAAGATGGTCAAGATGTTGTGAAGACGGG 276
QY 1271 tgtgagagt-----cagagtgaatggaagggcggtggttggatgagcgtgagattaa 1324
Db 275 GGTGAGGTGGATGATTAAGTGAATGTTGAGGAGGGATGATGGAAGCAGAGAGATTAG 216
QY 1325 gagggtgctcgcggaggtataaagagtcgag 1356
Db 215 GAAGTGTGGATGTGGTCATGGGGAGTGGAG 184

RESULT 12
AI779099      517 bp      mRNA      EST      29-JUN-1999
LOCUS      E1259978 tomato susceptible, Cornell Lycopersicon esculentum cDNA
DEFINITION      clone cLES719, mRNA sequence.
ACCESSION      AI779099
VERSION      AI779099.1 GI:5277140
KEYWORDS      EST.
SOURCE      tomato.
ORGANISM      Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 517)
D'Ascenzo,M., He,X., Lyman,J., Ronning,C.M., Craven,M.B., Fujii,C.Y.,
Liang,F., Upton,J., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
Giovannoni,J.J. and Martin,G.B.
Generation of ESTs from Pseudomonas susceptible tomato
Unpublished (1999)
On Jan 19, 1998 this sequence version replaced gi:2285531.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293

Email: dfrisch@CLEMSON.EDU
5 prime sequence.
Location/Qualifiers
1..517
/organism="Lycopersicon esculentum"
/cultivar="Ril-13 (Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="cLES719"
/clone_lib="tomato susceptible, Cornell"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOL"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLES - Tomato Pseudomonas Susceptible EST Library.
Directionally cloned cDNAs inserted into pBluescript.
SK(-) at 5' end with EcoRI and 3' end with XhoI site"

BASE COUNT      185 a   186 c   102 g   152 t
ORIGIN

Query Match      7.4%; Score 123; DB 44; Length 517;
Best Local Similarity 61.8%; Pred. No. 7.1e-21;
Matches 233; Conservative 0; Mismatches 135; Indels 9; Gaps 2;

QY 632 aatgaagaagaatgctgcacagcgtcgacaagaaccgaagcgaaggtcttgataaacac 591
Db 150 ATTCAAAGAGCAACTAGACATATTAGATGGTGAAGAAATCCAAAGGTACTTGTGAATAC 209
QY 692 gttcagcgcattagaaacgaacaaacgaacccatcgacaggtatgaactaatatccat 751
Db 210 ATTTGATGATTTAGAGCTAGAGCCACTCAAGCTATTGAAAATAACATTTTATTGGAAT 269
QY 752 cggccattatccccatcatcgatattctcagatggcaacgacccctcatcaagcaacaa 811
Db 270 TGGACCATTTGATTCCTTCATC---ATTCCTGGGTGGAAGA-----TTCATTGGAATC 320
QY 812 atctcagtgagagacctcttcagaagaacgcatgaaactacatgagcagctgctaaacac 871
Db 321 TTCATTGTGCTGATGCTCTTTTCAAAAGCTCAAAATGATGACTACATGAATGGTTAAACAC 380
QY 872 aaaaaccgaatcatcggtcggttacctggttcgttcggagagcctcctcgaggtcccgaaac 931
Db 381 AAAGCCTAAATCATCAATTTTATATCTCAATTTGGGAGTCTATTGAAATTTATCAAGAA 440
QY 932 ccaaatggaagaataagcaatagggtcttcagaccacacaaatcgccagttctctgggtgat 991
Db 441 CCAAAAGGAGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATCTTATGGGTAAT 500
QY 992 aagaagaacgaagagg 1008
Db 501 AAGAGATCAAGAAGAAG 517

RESULT 13
AW651280      658 bp      mRNA      EST      04-APR-2000
LOCUS      EST329734 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION      esculentum cDNA clone cLE116K15 5', mRNA sequence.
ACCESSION      AW651280
VERSION      AW651280.1 GI:7412518
KEYWORDS      EST.
SOURCE      tomato.
ORGANISM      Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (bases 1 to 658)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and
Tanksley,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
```


Job time: 62707 sec

[illegible]

BASE COUNT	149 a	128 c	86 g	174 t
ORIGIN				

Search completed: August 1, 2000, 14:38:35

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:50:22 ; Search time 47.63 Seconds
(without alignments)
237,706 Million cell updates/sec

Title: US-09-147-955-8

Perfect score: 2484

Sequence: 1 MVNKRHILLATFPAQGHINP.....GSSMNLKNFTRIINENAS 478

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2480	99.8	478	1 W92950	WO9905287 Seq ID 1
2	1473.5	59.3	461	1 W92949	WO9905287 Seq ID 9
3	1445	58.2	460	1 W92947	WO9905287 Seq ID 7
4	1422.5	56.9	443	1 W92948	WO9905287 Seq ID 8
5	1178.5	47.4	468	1 W92952	WO9905287 Seq ID 1
6	631	25.4	466	1 W92925	UDP-glucose:thiohy
7	631	25.4	471	1 W03756	Maize UDP-glucose
8	420	16.9	433	1 W13406	Solanum melongena
9	393.5	15.8	471	1 W98009	Medium-chain UDP-g
10	392.5	15.8	470	1 W47172	Glucosyl transfera
11	382	15.4	488	1 W64569	Potato solanidine
12	351.5	14.2	473	1 W56451	UDP-glucose:flavon
13	322	13.0	471	1 R49245	Anthocyanidin-3-gl
14	164.5	6.6	534	1 R26154	HUG-Br2. Isolated
15	156	6.3	533	1 R26153	HUG-Br1. Isolated
16	148.5	6.0	506	1 R10429	Ecdysteroid UDP-gl
17	145	5.8	535	1 W56750	Ecdysteroid UDP-gl
18	131.5	5.3	530	1 W47126	Uridine diphospho-
19	109.5	4.4	489	1 W20344	H. pylori putative
20	106.5	4.3	957	1 W20841	H. pylori putative
21	99.5	4.0	617	1 R15241	B. lautus endol cor
22	98	3.9	473	1 W64392	S. cerevisiae L947
23	98	3.9	1198	1 W64384	S. cerevisiae L947
24	97.5	3.9	339	1 Y07063	Renal cancer assoc
25	97.5	3.9	1006	1 W77290	Human differentiat
26	96.5	3.9	700	1 R13227	Novel endoglucanase
27	96	3.9	906	1 R42354	Human glutamate re
28	96	3.9	906	1 R84913	Human glutamate re
29	96	3.9	907	1 R11989	Glutamate receptor
30	94.5	3.8	414	1 R93982	Saccharothrix aero
31	94.5	3.8	865	1 R20670	Lipoxigenase. prod
32	94.5	3.8	865	1 R23797	Rice lipoxigenase.
33	91.5	3.7	74	1 R30165	UGT1 Exon 4 produc

34 90.5 3.6 1162 1 W19115 Murine long form O
35 89.5 3.6 574 1 Y00940 Aquifex VF-5 DNA p
36 89 3.6 2161 1 W87724 Rupestris stem pit
37 89 3.6 2710 1 R95016 C. difficile toxin
38 89 3.6 2710 1 W68387 Clostridium diffi
39 88.5 3.6 1015 1 W34259 Rat ob receptor is
40 88.5 3.6 1162 1 W24399 Rat ob receptor (f
41 88.5 3.6 1162 1 W23398 Rat ob receptor (w
42 88.5 3.6 1162 1 W34257 Rat wild-type ob r
43 88.5 3.6 1284 1 P81187 Sequence encoded b
44 88 3.5 409 1 R78525 Protein A-calmodul
45 87 3.5 1151 1 W77287 zebrafish differen

ALIGNMENTS

RESULT 1

W92950

ID W92950 standard; Protein; 478 AA.

AC W92950;

DT 14-MAY-1999 (first entry)

DE WO9905287 Seq ID 10.

KW Plant; flavonoid 5-transglycosylation activity; 576; variety; colour.

OS Torenia hyrira.

FH Key Location/Qualifiers

FT Misc_difference 64 /note= "Cys or Phe"

FT Misc_difference 65 /note= "Ser or Pro"

PN WO9905287-AL.

PD 04-FEB-1999.

PE 16-JUL-1998; J03199.

PR 25-JUL-1997; JP-200571.

PA (SUNR) SUNTORY LTD.

PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,

PI Yamazaki M;

DR N-PSDB; X02829.

DR WPI; 99-142940/12.

PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be

used to transform plants for improvement of plant coloration

PS Claim 2; Page 78-81; 89pp; Japanese.

CC This invention describes the isolation of plant proteins which have

flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used

in the production of plant varieties with improved colour, for example in

the production of flower varieties with desired coloration and cut

flowers harvested from them.

CC Sequence 478 AA;

SQ

Query Match

Best Local Similarity 99.8%; Score 2480; DB 1; Length 478;

Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVNKRHILLATFPAQGHINPSEPAKELLATGYVDQVTFSTVYALRRMRFTDPSSRID 60

DB 1 MVNKRHILLATFPAQGHINPSEPAKELLATGYVDQVTFSTVYALRRMRFTDPSSRID 60

QY 61 FVAXXDSYDDGLKGGDDGKNYMSMRKRGTALKDRTLKLNDAAMGSECYNRVSVFYSH 120

DB 61 FVAXXDSYDDGLKGGDDGKNYMSMRKRGTALKDRTLKLNDAAMGSECYNRVSVFYSH 120

QY 121 LFSWAAREVDVPSALLATEPATVDVYFVPCNGYADDIDAGSDQIQLNPLQSLKQD 180

DB 121 LFSWAAREVDVPSALLATEPATVDVYFVPCNGYADDIDAGSDQIQLNPLQSLKQD 180

QY 181 LPSFLPSSPARFRLTKKEKFDTLDKPEKAVLINTFDALETEQLKAIDRYELISIGPLI 240

DB 181 LPSFLPSSPARFRLTKKEKFDTLDKPEKAVLINTFDALETEQLKAIDRYELISIGPLI 240

QY 241 PSSIFSDGNDPSSSNKSVGGDLFRKADETVMDLNSKPSSVYVYSGSLLRLPKQWEE 300

DB 241 PSSIFSDGNDPSSSNKSVGGDLFRKADETVMDLNSKPSSVYVYSGSLLRLPKQWEE 300

QY 301 IAIGLSDTKSPVLWIRNREDEQBOAEKEEKLSTFFDRHGTERLGKIVTWCSQLDVLVT 360
DB 301 IAIGLSDTKSPVLWIRNREDEQBOAEKEEKLSTFFDRHGTERLGKIVTWCSQLDVLVT 360
QY 361 HKSVCFCVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKWIEDVWRSQVVRVNEEGV 420
DB 361 HKSVCFCVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKWIEDVWRSQVVRVNEEGV 420
QY 421 VDRRETKRCVSVIKSRELRESAMWKLAKEMDEERGSSMNNLKNFTIRINENAS 478
DB 421 VDRRETKRCVSVIKSRELRESAMWKLAKEMDEERGSSMNNLKNFTIRINENAS 478
RESULT 2
W92949
ID W92949 standard; Protein; 461 AA.
AC W92949; AC
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 9.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety: colour.
OS Verbena hybrida.
PN W0905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR N-PSDB; X02828.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Claim 2; Page 76-78; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
SQ Sequence 461 AA;

Query Match 59.38; Score 1473.5; DB 1; Length 461;
Best Local Similarity 61.5%; Pred. No. 8.8e-137;
Matches 295; Conservative 71; Mismatches 85; Indels 29; Gaps 11;
QY 2 VNKRHILLATPAQGHINPSLEFAKRLNTGYVDQVTFYTSVVALRRM-RFETDPSRSD 60
DB 1 MSRAHVLLATPAQGHINPALQFAKRLNADI--QVTFYTSVVALRRMSRTAAGSNGLIN 58
QY 61 FVAXXSDYDGLKGGDKGNVSEMRKGTAKLKDILKINDAAMGSECYNRVSVFYVSH 120
DB 59 FVSFDGYDDGQPGDGNVSEMRKGTAKLKDILKINDAAMGSECYNRVSVFYVSH 114
QY 121 LFSWAAEAREVDVPSALLWTEPATVDFVYFNGVADDIDAGSDQIQILP-NLPOLSKQ 179
DB 115 LFAWAAKVAAREPHLSALLWTEPATVDFVYFNGVADDIDAGSDQIQILP-LPGSLPVLQ 174
QY 180 DLPSFLLPSPARFRTLMKEFDTLDEPKAKVLINTFDALTEQLKADRYELISIGPL 239
DB 175 DLPSFLLPSPARFRTLMKEFDTLDEPKAKVLINTFDALTEQLKADRYELISIGPL 234
QY 240 IPSFISDGNDDSSNKSNGYGGDLFRKA--DETMDLWNSKPSSVYVYFSGSLRLLPKPQ 297
DB 235 IPSA-FLDGKDP--SDRSFGDLPEKSGNDDCLLEWLSNPRSSVYVYFSGSFVNTKSO 291
QY 298 MEETAGLSDTKSPVLWIRNREDEQBOAEKEEKLSTFFDRHGTERLGKIVTWCSQLD 357
DB 292 MEETAGLSDTKSPVLWIRNREDEQBOAEKEEKLSTFFDRHGTERLGKIVTWCSQLD 340
QY 358 VLTHKSVGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKWIEDVWRSQVVRVNEE 417
DB 341 VLTHPSLACGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKWIEDVWRSQVVRVNEE 400

QY 418 GGVDRRETKRCVSEVI----KSRELRESAMWKLAKEMDEERGSSMNNLKNFTIRI 473
DB 401 GSVYDGEIRRCIEEYMDGGEKSRKRESAGKWKDLARKAMEED-GSSVNNLKVFLDEVV 459
RESULT 3
W92947
ID W92947 standard; Protein; 460 AA.
AC W92947; AC
DT 14-MAY-1999 (first entry)
DE W0905287 Seq ID 7.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety: colour.
OS Perilla frutescens.
PN W0905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR N-PSDB; X02826.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Claim 2; Page 71-73; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
SQ Sequence 460 AA;

Query Match 58.28; Score 1445; DB 1; Length 460;
Best Local Similarity 61.1%; Pred. No. 5.6e-134;
Matches 292; Conservative 65; Mismatches 89; Indels 32; Gaps 11;
QY 4 KRHILLATPAQGHINPSLEFAKRLNTGYVDQVTFYTSVVALRRMRFETDPS--RI 59
DB 3 RRVLLATPAQGHINPALQFAKRLKAG--TDVTFYTSVVALRRMANTASAAGPPGL 60
QY 60 DFVAXXSDYDGLKGGDKGNVSEMRKGTAKLKDILKINDAAMGSECYNRVSVFYVYS 119
DB 61 DFVAFSDGYDDGDLKPGCGKRYMSEMKARGSEALRNLLNHHID-----VTFVYYS 110
QY 120 HLFSAWAAEAREVDVPSALLWTEPATVDFVYFNGVADDIDAGSDQIQILP-NLPOLSKQ 179
DB 111 HLFSAWAAEAREVDVPSALLWTEPATVDFVYFNGVADDIDAGSDQIQILP-NLPOLSKQ 170
QY 180 DLPSFLLPSPARFRTLMKEFDTLDEPKAKVLINTFDALTEQLKADRYELISIGPL 239
DB 171 SLPTFLLPSPARFRTLMKEFDTLDEPKAKVLINTFDALTEQLKADRYELISIGPL 230
QY 240 IPSFISDGNDDSSNKSNGYGGDLFRKADE--TYMDLWNSKPSSVYVYFSGSLRLLPKPOM 298
DB 231 IPSA-FLDGKDPSET--SYGGDLFEKSENNCEVMDLTKPKSSVYVYFSGSLRFPKQM 287
QY 299 EETAIGLSDTKSPVLWIRNREDEQBOAEKEEKLSTFFDRHGTERLGKIVTWCSQLD 358
DB 288 EETAGLSDTKSPVLWIRNREDEQBOAEKEEKLSTFFDRHGTERLGKIVTWCSQLD 341
QY 359 LTHKSVGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKWIEDVWRSQVVRVNEEG 418
DB 342 LAHPALGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKWIEDVWRSQVVRVNEEG 401
QY 419 GGVDRRETKRCVSEVI----KSRELRESAMWKLAKEMDEERGSSMNNLKNFTIRI 472
DB 402 G-VDSGEIERCEVMDGGEKSKLVRENAIKNKTFLAREMGED-GSSLKNLNAFLHQV 457

RESULT 4
W92948

ID W92948 standard; Protein; 443 AA.
 AC W92948; 1999 (first entry)
 DE W09905287 Seq ID 8.
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
 OS Perilla frutescens.
 PN W09905287-A1.
 PD 04-FEB-1999.
 PP 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNR) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 PI Yamazaki M;
 DR WPI: 99-142940/12.
 DR N-PSDB: X02827.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 PT used to transform plants for improvement of plant coloration
 PS Claim 2; Page 74-76; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have
 CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
 CC in the production of plant varieties with improved colour, for example in
 CC the production of flower varieties with desired coloration and cut
 CC flowers harvested from them.
 SQ Sequence 443 AA;

 Query Match 56.9%; Score 1412.5; DB 1; Length 443;
 Best Local Similarity 61.3%; Pred. No. 8.5e-131;
 Matches 282; Conservative 66; Mismatches 83; Indels 29; Gaps 10;

 Qy 4 KRHTLATFPAQGHINPSLEFAKRLNTGYDQVTFVVALRRMRFETDPSS-RI 59
 Db 3 RRRVLLATFPAQGHINPALQFARLKAG--TDVTFVTVARWAMTASAGNPPGL 60

 Qy 60 DFVAXSDYDGLKGGDKKNYSEMKGRTKALKDITLKLNDAAMGSECYNRVSVFYV 119
 Db 61 DFVAFSDYDGLKPGDGKRYNEMKARGSEALRNLLNND-----VTFVYV 110

 Qy 120 HLFSAEAAREVDVPSALLWIEPATVDFVYFYNGYADDTDAGSDIOIQLPNLPOLSKQ 179
 Db 111 HLFSAEAAREVLSVFTALLWIEPATVLCIYFYNGYADEIDAGSNEIQPLPSLEQR 170

 Qy 180 DLPSFLPSPAPRTLMKEKFTDLKPEKAKVLINTDALETKQAKIDRYELISGTL 239
 Db 171 SLPTFLPATPERFLMKKEKLETLDEGEKAKVLNTPDALEPDALTADRYELIGIPL 230

 Qy 240 IPSISFGDNDPSSNKSNGYDGLFRKADE--TYMDWLNKPSVYVYVSGSLRLKLPQM 298
 Db 231 IPSA-FLDGEDPSET--SYGCDLPESENNCVENWLNKPSVYVYVSGSLRFPKQAM 287

 Qy 299 BEIAGISDTSKPSVLVIRNEEGDQEQAEKEELLSFFDRHGRTERLGKIVTWCOLDY 358
 Db 288 BEIGKGLLAGCRPLWIRBQKNDGDEEEEEE--LSCIGE--LKKMGKIVSWCSOLEV 343

 Qy 359 LTHKSVGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKMIEDVWRSVYRVN 418
 Db 344 LAHPALGCFVTHCGWNSAIESLSCGTPVAVFQWFDQGTNAKLIEDAWGTGVRVNM 403

 Qy 419 GVDRREIKRCVSEI-----KSRELRESAMMMKGLAKEAM 454
 Db 404 G-VDCETERCVEMVMDGDKTLVRENAIKWTLARQAM 442

 RESULT 5
 W92952
 ID W92952 standard; Protein; 468 AA.
 AC W92952;
 DT 14-MAY-1999 (first entry)
 DE W09905287 Seq ID 12.
 KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
 OS Petunia hybrida.
 PN W09905287-A1.
 PD 04-FEB-1999.

PF 16-JUL-1998; J03199.
 PR 25-JUL-1997; JP-200571.
 PA (SUNR) SUNTORY LTD.
 PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
 PI Yamazaki M;
 DR WPI: 99-142940/12.
 DR N-PSDB: X02831.
 PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
 PT used to transform plants for improvement of plant coloration
 PS Claim 2; Page 83-85; 89pp; Japanese.
 CC This invention describes the isolation of plant proteins which have
 CC flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
 CC in the production of plant varieties with improved colour, for example in
 CC the production of flower varieties with desired coloration and cut
 CC flowers harvested from them.
 SQ Sequence 468 AA;

 Query Match 47.4%; Score 1178.5; DB 1; Length 468;
 Best Local Similarity 51.7%; Pred. No. 1.1e-107;
 Matches 247; Conservative 74; Mismatches 122; Indels 35; Gaps 14;

 Qy 6 HILLATFPAQGHINPSLEFAKRLNTGYDQVTFVVALRRMRFET--DPSSRIDPVA 63
 Db 5 HVILTFPAQGHINPALQFARLNKMGII--EVTFTSIYAQSMDEKSLINAPKGLNFI 62

 Qy 64 XDSYDGLKGGDKKNYSEMKGRTKALKDITLKLNDAAMGSECYNRVSVFYVSHLFS 123
 Db 63 FSDGDFGDFHSDKDPVYMSQLKCGSETVKKIILTCSENGQPIVC-----LLYSIFLP 116

 Qy 124 WAAEVAAREVDVPSALLWIEPATVDFVYFYNGY-----ADDIDAGSDIOIQLPNLPOLSKQ 179
 Db 117 WAAEVAAREVHVPISALLWISQPATILDIYFNFHGYEKAMANSNDPNWSIQLPGLLETR 176

 Qy 180 DLPSFLP--SSPAPRTLM---KEKFTDLKPEKAKVLINTDALETKQAKIDRYELIS 235
 Db 177 DLPSFLPYGAKGSLRVALPFPKELITDLDAETTPKILVNTDELEPEALNAIEGYKFG 236

 Qy 236 IGPLPISIESDNDPSSNKSNGYDGLFRKADETYMDWLNKPSVYVYVSGSLRLPK 295
 Db 237 IGPLPSA-FLGNDPLDA--SFGDGLFQNSND--YMWLNKPSVYVYVSGSLRNPSI 292

 Qy 296 POMBEIAGISDTSKPSVLVIRNEEGDQEQAEKEELLSFFDRHGRTERLGKIVTWCQ 355
 Db 293 SQMBEISKGLDIDIGRPLWIKENEG-----KEBENKKGICIEE--LEKIGKIVPWCQ 345

 Qy 356 LDVTHKSVGCFVTHCGWNSAIESLACGVPVVCFFQWFDQGTNAKMIEDVWRSVYRVN 415
 Db 346 LEVLKHPISLGCFSHCGWNSAIESLACGVPVVPVAFPOWTDQMTNAKQVEDVWKSQVYRVIN 405

 Qy 416 EEGGVDRREIKRCVSEI-----KSRELRESAMMMKGLAKEAMDEERGSSNNLNKFI 469
 Db 406 BD-GVVESEIKRCIELVMDGGEKGBELRNKAKKWLAREAV--KEGSSSHKLNKAFI 461

 RESULT 6
 W09825
 ID W09825 standard; Protein; 466 AA.
 AC W09825;
 DT 15-JUL-1997 (first entry)
 DE UDP-glucose:thiohydroximate S-glucosyltransferase.
 KW Glucosinolate; UDP-glucose:thiohydroximate S-glucosyltransferase;
 KW S-GT; transgenic plant; rapeseed oil; oilseed rape; canola.
 OS Brassica napus cv. Westar.
 FH Key location/Qualifiers
 FT misc_difference 2
 FT /note- "residue 2 is Val in other S-GT isoforms"
 FT misc_difference 10..11
 FT /note- "a Lys residue is inserted between amino
 FT acids 10 and 11 in some S-GT isoforms"
 FT misc_difference 12
 FT /note- "residue 12 is Ser in some S-GT isoforms"
 FT misc_difference 43

PA (RSE) RESEARCH CORP TECHNOLOGIES INC.
 PI Bandurski RS, Szczeglowski K, Szerszen JB;
 DR WPI; 96-068875/07.
 DR N-PSDB; T10106.
 PT UDP-glucose indol-3-yl:acetyl:glucosyl transferase DNA and protein
 PT - used in sense or anti-sense orientation for the genetic control of
 PT plant IAA levels and plant growth
 PS Claim 3; Page 21-23; 41pp; English.
 CC This sequence encodes a UDP-glucose-indol-3-
 CC ylacetylglucosyltransferase (IAGlu-transferase) from maize (Zea mays).
 CC Over-expression of this sequence in transgenic plants will result
 CC in at least 1 of the following effects: inhibition of apical
 CC dominance; inhibition of stem elongation; inhibition of cell
 CC enlargement and increased numbers of stems per plant as compared
 CC with a wild-type plant. The ability to control the expression of
 CC this enzyme allows the control of free indoleacetic acid levels in
 CC plants, thereby affecting plant growth rates. Antisense constructs
 CC may be used to inhibit synthesis of this enzyme.
 SQ Sequence 471 AA;

Query Match 25.4%; Score 631; DB 1; Length 471;
 Best Local Similarity 33.7%; Pred. No. 9.4e-54;
 Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

QY 6 HLLATFPAQGHINPSLEFAKRLNTGYVDVTFSTVSVALRRMRFETDPSRIDFVAXX 65
 DB 4 HVLVVPFGQGHMMPVQFAKRLASKGVA--TTLVTRFIQTADVADHPAM---VEAIS 58
 QY 66 DSYDGLKKGDDGKNYSEMRKGTALKDTLIKLNDAAMGSECYNRVSVVYSHLFSWA 125
 DB 59 DGHDEG---GFASAAGVAEYLEKQAAASASLASLVEARASSA--DAFTCVYDSDVDM 113
 QY 126 AEVAREVDVPSALLWIEPAVDFVYFYFNG-----YADDIDAGSDQIQLP----NL 173
 DB 114 LPVARMGLPAVPSTQSCAVSYVHFSGRLAVPPGAADGSDGAGAAALSEAFUGL 173
 QY 174 POLSKQDLPFLSPSPAREFTLMKEKFDLDKEPRAK-VLINTFDATEOLKAIDRY- 231
 DB 174 PEMERSELPSEVFDHGP--YPTIAMQAIKQFAHAGKDDWLVFNFSPEELETEVLACLTXYL 231
 QY 232 ELISIGPLIPSSIFSDGNDPSSNK-SYGGDLERKADEYTMWLNSKPESSVYVYSGSL 290
 DB 232 KARATGCPVP--LPTAGTAGANGRIYGANLV-KPEDACTKWLDTKPDORSVAYVSGSL 288
 QY 291 LALPQMEETAIGLSDTKSPVLWVIRNNEEGDEQEAEEELKLSFFDRHCTER-LGKI 349
 DB 289 ASLIGNAQKEELARGLLAAGKFFLWVR-----ASDEHQVPRYLLAEATATGAAMV 338
 QY 350 VTWCSQLDLVTHKSVGCFVTHCGNSAIESLACGVPVPCFQWFOGNTNAKMIEDVWRSG 409
 DB 339 VPWCQQLDLAHPAVGCFVTHCGNSLTLEALSGVPMVAMALWTDQPTNARNVELANGAG 398
 QY 410 VVRVNEEGGVVDRREIKRCVSEVKSRE----LRSAAMWKGAKAEAMDEBERGSSMNL 465
 DB 399 VRAREDACAGVFLRGVERCVRAVMDGGEAASAAKAKAGENWRDRAAV-APGSSDRNL 457
 QY 466 KNFI 469
 DB 458 DEFV 461

RESULT 8
 W13406
 ID W13406 standard; Protein; 433 AA.
 AC W13406;
 DT 10-JUN-1997 (first entry)
 DE Solanum melongena flavanoid-3-glucosyl transferase.
 KW Eggplant; flavanoid-3-glucosyl; transferase; glucose; flavonoid;
 KW pigment; production; seed; plant.
 OS Solanum melongena.
 PN J09056385-A.
 PD 04-MAR-1997.

PF 25-AUG-1995; 238943.
 PR 25-AUG-1995; JP-238943.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI; 97-206631/19.
 DR N-PSDB; T62123.
 PT Solanum melongena flavonoid-3-glucosyl transferase and DNA encoding
 PT it - catalyses transfer of glucose to 3-position of flavonoid
 PT pigment in plants
 PS Claim 4; Pages 9-10; 14pp; Japanese.
 CC The present sequence is the Solanum melongena (eggplant)
 CC flavanoid-3-glucosyl transferase (F3GT), which catalyses the
 CC transfer of glucose to the 3-position of flavonoid pigments. The
 CC F3GT DNA is useful for the production of the seed of a plant. A
 CC S. melongena seed was cultured under a red light and a flavanoid
 CC derived. A cDNA library in which the pigment was concentrated was
 CC prepared. A flavanoid glucosidising enzyme gene was cloned and its
 CC structure and mRNA expression pattern analysed. The clone was
 CC expressed in E. coli.
 SQ Sequence 433 AA;

Query Match 16.9%; Score 420; DB 1; Length 433;
 Best Local Similarity 27.1%; Pred. No. 5.1e-33;
 Matches 136; Conservative 93; Mismatches 161; Indels 112; Gaps 26;

QY 6 HLLATFPAQGHINPSLEFAKRLNTGYVDVTFSTVSVALRRMRFETDPS-----SRID 60
 DB 7 HIAFLAFLPFGTHATPLTLVQKI--SPFLPSTIFSP-----FNTSSNSSIFSKVP 56
 QY 61 FYVAXDSYD--DGLKKGDDGKNYSEMRKGTALK-----DTLI--KLNDAAAGSECYNR 112
 DB 57 NOENIKIYNVWDGKEND-----TPFGLAIAKLFIOSTLLISKITEA---EEETG 105
 QY 113 VSF-VVYSHLFSWA--AEVAREVDVPSALLWIEPAVDFVYFYFNGYADDIDAGSD-QI 168
 DB 106 VKFSCIFSDAFMCLFVKLPKKNAPGAVYMTGGSCSLAVHL-----YTLIRSKETSL 160
 QY 169 QLPNL-----POLSKQDLPFLSPSPARFR-TLMKEKFDLDKEPRAKVLINTF 217
 DB 161 KIPGFSSTLSINDIPPEVTAEDLEG--PMSSMLYNALNLHKAADA-----VVLNSF 209
 QY 218 DALETEQL--KAIDR--VELISIGPLIPSSIFSDGNDPSSNKSXGGDLFRKADET-YMD 272
 DB 210 QELDRDPLINKDLQKNQKVENIGLVLQS-----SRKLDESGCIQ 250
 QY 273 WLNSKPSSVYVSGSLRLPKPQMEETAIGLSDTKSPVLWVIRNNEEGDEQEAEEEE 332
 DB 251 WLDQKESVYVLSGCTVTLPPNEIGSIAEALETKTPTFINSLRNNGVKNLPK----- 304
 QY 333 KLLSPFDRHGTERTLGKIVTWCSQLDLVTHKSVGCFVTHCGNSAIESLACGVPVVCFPQW 392
 DB 305 ---GFELR--TFEFGKIVSWAPQLEILAHKSVGVFVTHCGNSILEGIFSGVPMICRPFF 359
 QY 393 FQGGTNAKMIEDVWRSGVVRVIRNNEEGGVDRREIKRCVSEVI---KSRELRESAMMWKGL 449
 DB 360 GQKLNRSRVESVWEIGLOI---EGGIFTKSGIISALDTFFNEEKGKILRENV---EGL 412
 QY 450 AKEAMDEERGSSMNLKNFTR 471
 DB 413 KEKAL--EAVNQMEVQOKISR 432

RESULT 9
 W98009
 ID W98009 standard; Protein; 471 AA.
 AC W98009;
 DT 21-JUN-1999 (first entry)
 DE Medium-chain UDP-glucose:fatty acid glucosyltransferase.
 KW UDP-glucose:fatty acid glucosyltransferase; tomato;
 KW 1-O-acyl-a-glucose; acyl donor; esterification.
 OS Lycopersicon pennellii
 OS Lycopersicon esculentum.
 PN W09909144-AI.


```

Db 374 MYTWPFVFAEQFENEKLYTEVMRSCAGYGSQKWKTASEG--VKREAIAKAIKRVNASEET 431
QY 439 --LRESAMRWGLAKEMDERGSSMNNLKNFTIRI 472
Db 432 EGRFRRAKEYKEMAREAI-BEGGSSYNGWATLIQDI 466

RESULT 11
W64569
ID W64569 standard; Protein; 488 AA.
AC W64569;
DE Potato solanidine UDP-glucose glucosyltransferase.
KW Solanidine UDP-glucose glucosyltransferase; SGT; potato; tomato;
KW glycoalkaloid; solanidine; solanaceous plant; transgenic plant.
OS Solanum tuberosum cv. Lemhi Russet.
FH Key Location/Qualifiers
FT Region 12..46 "conserved domain"
FT Region 110..143
FT Region /note="putative steroid binding region"
FT Region 351..401
FT Region /note="UDP-glucose binding region"
PN W09834471-A1.
PD 13-AUG-1998.
PF 30-JAN-1998; U01864.
PR 07-FEB-1997; US-797226.
PA (USDA ) US SEC OF AGRIC.
PI Allen PV, Belknap W, Friedman M, Moehs CP, Rockhold DR,
PI Stapleton A;
DR WPI: 98-446828/38.
DR N-PSDB: V49609.
PT New DNA encoding solanidine UDP-glucose glucosyl-transferase and
PT related vectors - and transgenic plants, used to reduce contents of
PT steroidal glyco-alkaloid(s) in plants, specifically potato
PS Example 1; Fig 2; 54pp; English.
CC This is the amino acid sequence of potato cv. Lemhi Russett
CC solanidine UDP-glucose glucosyltransferase (SGT). It was deduced
CC from an SGT cDNA sequence (see V49609). SGT is involved in the
CC biosynthesis of steroidal glycoalkaloids in solanaceous plants,
CC catalysing the UDP-glucose dependent conversion of the aglycone
CC solanidine to gamma-chaconine. The invention relates to DNA
CC sequences which encode SGT, and their use, particularly use of an
CC antisense DNA construct to inhibit SGT activity and glycoalkaloid
CC levels in solanaceous plants. Transgenic plants are claimed,
CC particularly tomato and potato, that have reduced contents of toxic
CC steroidal glycoalkaloids. The inactivation of glycoalkaloid
CC biosynthetic pathways is beneficial to reduce or eliminate
CC glycoalkaloid biosynthesis during storage and shipping.
CC Sequence 488 AA;

Query Match 15.4%; Score 382; DB 1; Length 488;
Best Local Similarity 23.5%; Pred. No. 3.4e-29;
Matches 127; Conservative 92; Mismatches 186; Indels 136; Gaps 18;

QY 6 HILLATFPAQGHINPSLEFAKRLNTGYDQVTFVTSVVALRMRFETDPSRID----- 60
Db 12 HWLFLPFSAGHFPLVNAARLFRSGV--KATILTPHALLFRSTIDDDVRISGFPTIS 69
QY 61 -EVAXXDSDYDGLKGGDDGKNYMS--EMRXR-----GTKALKDPLIKLNDAAAGSECY 110
Db 70 IVTIKPPSAEVLPGIEFSNATSPENPKIFVALSLLQKPMEDIKIRELPDCIFSDMY 129
QY 111 NRVSFVYSHLFSWAAEVARVDVPSALLWTEPATVDYVYFYFNGYADDDIAGSDQQL 170
Db 130 -----FPWVTDIADLHPRIYLNLSAYMCYSINH-----NLKVR 165
QY 171 PNLQSLKQDLPSSFLLSPSPARFTLMKEKFDITDKPEKAK----- 211
Db 166 PH-KQNLDESQSFVVGPLPEIKFLSQLTDDLKSDQKTVFDELLQVEDSEERSYG 224
QY 212 VLINTFDALETQLKADRYELIS-----IGPLI-----PSSIFSDGNDPSSNSKY 258

```

```

Db 225 IVHDTFYELPAY---VDYOKLAKPKCWHEGPLSHFASKIRSKELISEHNN----- 273
QY 259 GGDLPFRKADETYMDLNSKPSSVYVYFSGSLRLRLPKPOMEETAIGLSDTKSPVLWVIRR 318
Db 274 -----NEIVIDLWLNAAQPKSVLYVYFSGSMARFPESQLENTAQALDASNPVFIPLRP 325
QY 319 NEE-----GDEQSOAEERKLLSFFDRHCTERLGLKIVTCWSQOLDVLTHKSGVCPVTHC 371
Db 326 NEETASWLPVGNLEDR-----TKGLYIKGWVLPQLTIMEHSATGGPPTH 370
QY 372 GWSAIESLACGVPVVCPPQWFDQGTNAKME-----DWRSGVRVRVNERGGV 421
Db 371 GTNSVLEALTFCVPMITWPLYADOFYNEKVVYRGLGKIGIDWNNEGIEI-----TGPVI 426
QY 422 DRRETKRCVSEVIKSR-----ELRESAMWKKGLAKEMDERGSSMNNLKNFTIRINE 475
Db 427 ESAKREAIERLWISNGSEEIINIRVMAKSMQAQNTN-EGGSWNNLTALIQIKNY 485
QY 476 N 476
Db 486 N 486

RESULT 12
W56451
ID W56451 standard; Protein; 473 AA.
AC W56451;
DE UDP-glucose:flavonoid 3,5-O-glucosyl transferase amino acid sequence.
KW UDP-glucose:flavonoid 3,5-O-glucosyl transferase; gentian; rose plant;
KW blue flower.
OS Unidentified.
PN J10113184-A.
PD 06-MAY-1998.
PR 15-OCT-1996; JP-272253.
PA (TOLG ) TOKYO GAS CO LTD.
DR WPI: 98-315473/28.
DR N-PSDB: V23108.
PT Gentian flavonoid glucosyl transferase gene - used to produce
PT recombinant rose plants with blue petals.
PS Claim 1; Pages 6-7; 10pp; Japanese.
CC The present sequence represents a UDP-glucose:flavonoid 3,5-O-glucosyl
CC transferase enzyme. The nucleic acid sequence was amplified from
CC RNA which had been isolated from gentian (sic) petals. The gene is
CC used to transform rose plants to produce blue flowers.
CC Sequence 473 AA;

Query Match 14.2%; Score 351.5; DB 1; Length 473;
Best Local Similarity 25.3%; Pred. No. 3.3e-26;
Matches 133; Conservative 89; Mismatches 194; Indels 109; Gaps 21;

QY 2 VNKRHLLATFPAQGHINPSLEFAKRL-----LNTGYDQVTFVTSVVALRMRFETDP 55
Db 1 MKKAEVIIPFGISHLGSTVELAKLAERNEHLSISVI--IKFPNDKVSNNLSLST 58
QY 56 SSRIDFVAXXSDYDGLKKG-----DDGKNYMSMRKRGTKALKDPLIKLNDAAAGSECY 111
Db 59 ASRIKVI-----LKQETISITETGPTFKQFIESHKAQVRDL-----AGVSACES 104
QY 112 -RVSFVYSHLFSWAAEVARVDVPSALLWTEPATVDYVYFYFNGYAD-----DIDAG 164
Db 105 VELSGVVIDMCTSMIDVANEFEVPSYVFTSSAMLGL-WFHQSLRDNFGKYVDIKOS 163
QY 165 SDQIQLPNLQSLKQD-LPSSLFPSPARFTLMKEKFDITD-----KEPKAK 211
Db 164 ETVLSIPAFONLVPYGVLPMEI-----FNTEDGCDASLOIGKRFRTKG- 207
QY 212 VLINTFDALETQLKADRYELI---SIGPLIPSSIFSDGNDPSSNSKYSGDLFRKAD 267
Db 208 IIINTFLESHAIESLSTDETIPPVYVGPILGPX-----GSSIESLET 252

```

QY 268 ETYMDLNSKPSSVYVYFSGSLRLPKPQMBEIAIGLSDTKSPVLWVIR-----NERG 322
 Db 253 EKLKWLDMQPEKSVYVYFCGSLGHFGAQKEIAYALEGSGHFLMSLRKPLPLGKFE 312
 QY 323 DEQOAEKEELKSFDFRHRGTERLGIKIVTWCSQLDVLTHKSGVCFVTHCGWNSAIESLAC 382
 Db 313 PGEYENLEEVLEPEGLER--TANTGWVIGWAPQTAVALSHNAVGFVSHCGWNSLLESIW 370
 QY 383 GVPVCFQFQMGCTNAXMIEDVNRSGVVRV-----NREGGVYDREIKRCVSEV 433
 Db 371 GVPVATWPLFABQOMNA--FELVKELGLAVEVKMDYKDKYKNPDADEIVRADV---IBEK 425
 QY 434 IK-----SRELSEAMWKLAKEMDERGSSMNNLNKFNITRII 473
 Db 426 IKILMPENGIRKKYKENKESRLAV-REGGSSSASLKDFINDVI 469

RESULT 13

R49245
 ID R49245 standard; Protein: 471 AA.
 AC R49245;
 DT 08-AUG-1994 (first entry)
 DE Anthocyanidin-3-glucoside rhamnosyltransferase (3RT).
 KW Anthocyanidin-3-glucoside rhamnosyltransferase;
 KW glucosyltransferase; inflorescence; flowering plants;
 KW transgenic plant; Petunia hybrida.
 OS Petunia hybrida.
 PN W09403591-A.
 PD 17-FEB-1994.
 PR 30-JUL-1993; AU-003846.
 PR 30-JUL-1993; AU-003846.
 PA (ITEL-) INT FLOWER DEV PTY LTD.
 PI Brugliera F. Holton TA.
 DR WPI: 94-065680/08.
 DR N-PSDB; 056241.
 PT Nucleic acid encoding glycosyltransferase enzymes - used for
 PT producing transgenic plants with altered inflorescence properties
 PT including modified petal colours
 PS Claim 24; Page 52-55; 76pp; English.
 CC The DNA encoding the anthocyanidin-3-glucoside rhamnosyltransferase
 CC is used in the construction of a vector which can then be used
 CC to transform plants. The transgenic plants are then capable of
 CC expressing the anthocyanidin-3-glucoside rhamnosyltransferase which
 CC confers altered inflorescence properties including modified petal
 CC colours.
 SQ Sequence 471 AA;

Query Match 13.0%; Score 322; DB 1; Length 471;
 Best Local Similarity 24.3%; Pred. No. 2.6e-23;
 Matches 120; Conservative 81; Mismatches 218; Indels 74; Gaps 16;

QY 6 HILLATFPAQGHINPSLEFAKRLNTGYVDQVTFSTSVYALRRMRETFDPSRIDFVAXX 65
 Db 11 HVVMFFFAFGHISPPVOLANKL--SSYGVKVSFTASGNASRVKSLNSAPTTHIVPLT 68
 QY 66 DSYDDGLKGDGCKNYMSEMRRKGTALKDTLKLINDAANGSECYNRVSVYSHLFSNA 125
 Db 69 LPHVEGLPPGABS---TAEITPASAEHLKVALDLMQPOIKTLHLKPHVLEDFDAQEWL 125
 QY 126 AEYAREVDVPSALLWIEPATVDVYFYFYNGYADD-IDAGSDQIQFLNLPOLSKDLPSPF 184
 Db 126 PKNANGLGKTI-----VYYSVVVALSTAELTCTPARVLEPKKPSLEDMKKPPL 173
 QY 185 LLP-SSPARFRILMKEF-----DTLDREPK--AKVLTINTDALETOLKA 227
 Db 174 GFPQTSVTSVTFEARDFLVYPKSPHNGPTLYDRIQSLRGCSAILAKTCSQMEGYIKY 233
 QY 228 ID---RYELISICPLIPSSIFSDGCKNYSNKSXGGDLFRKADETYMNLNKPSSVYV 284
 Db 234 VEAQKNKPVFLIGPVVP-----DPPSG-----KLEEKWATLNLKEGGTVIY 275

QY 285 VSTGSLRLPKPQMBEIAIGLSDTKSPVLWVIRNEEGDEQOAEKEELKSLSPFDRHGTE 344
 Db 276 CSFGSETFLTDQVKELALGLTGTPFLVL--NFPANVDVSAELNRLPEGF----LE 329
 QY 345 RL---QKIVT-WCSQLDVLTHKSGVCFVTHCGWNSAIESLACGVPVYCFQFQMGDOGTNAK 400
 Db 330 RVKDKGIHSGWVQOQHILAHSSVGCYCHAGFSVIEALVNDQVVLQKGDQITLNAK 389
 QY 401 MIEDVNRSGVVRVNEEGGVYDREIKRCVSEVTKSRE-----LRESAMWKLAKEM 452
 Db 390 LVSGDMEAGVNEINRDEDDYFGKEDIKEAVEKVMYDVKEPKGLIRENQKWEFLINKD 449
 QY 453 AMDEERGSSMNNL 465
 Db 450 IQSKYIGNLVNEM 462

RESULT 14

R26154
 ID R26154 standard; Protein: 534 AA.
 AC R26154;
 DT 27-JAN-1993 (first entry)
 DE HUG-Br2.
 KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
 KW monoglucuronide; diglucuronide.
 OS Homo sapiens.
 FH Key
 FT region Location/Qualifiers
 FT 12..22 /note= "putative membrane-insertion signal"
 FT region 492..508 /note= "putative membrane-anchoring peptide"
 FT modified_site 348 /note= "predicted Asn-linked glycosylation site"
 FT misc_difference 282..285 /note= "residues encoded by TCCCAACGGGAAG !"
 PN W09212987-A.
 PD 06-AUG-1992.
 PR 10-JAN-1992; U00282.
 PR 10-JAN-1991; US-639453.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Owens IS, Rittler JK;
 DR WPI: 92-284593/34.
 DR N-PSDB; Q27369.
 PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
 PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
 PT types I and II
 PS Disclosure; Fig 9A-I; 99pp; English.
 CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
 CC been isolated. They are referred to as HUGBr1 (Q27369) and HUGBr2
 CC (Q27370) (Rittler et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
 CC upon expression individually in COS-1 cells, encode isoforms that
 CC catalyse the formation of the two bilirubin monoglucuronides and
 CC the diglucuronide.
 CC The cDNAs contain identical 3' ends (1469 bp in length) to each
 CC other and to that of the human phenol transferase cDNA, HLUGP1
 CC (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).
 CC In contrast, they have unique 5' ends.
 SQ Sequence 534 AA;

Query Match 6.6%; Score 164.5; DB 1; Length 534;
 Best Local Similarity 20.5%; Pred. No. 1.1e-07;
 Matches 99; Conservative 80; Mismatches 175; Indels 129; Gaps 22;

QY 14 AQGH---INPSLEFAKRLNTGYVDQVTFSTSVYALRRMRETFDPSRIDFVAXXDSY 68
 Db 53 ARGHQAVALTPGREGYAHQ-----EKEFTLTAYAV-----PWTQKEFDRVTGLY 96
 QY 69 DDG-----LKKGGDGKNYMSMKRGTALKDTLKL--NDAAMGSECYNRVSV--VV 117
 Db 97 TOGFETEHELLKRYSRMAIMNV---SLALHRCCEVLLHNEALI--RHNLNATSFQVVL 150
 QY 118 YSHLFSMAEAREVDVPSALLWIEPATVDVYFYFYNGYADDIDAGSDQIQFLPN--LPQ 175

Db 151 TDPVNLGGAVLAKYLSIPAVFW-----RYPCDLDFKQGCNPSPSYTPK 196
QY 176 LSK-----QDLPSFLP-----SSPFRFTLMKEKFDTLDKPEPAKVLINTFD 218
Db 197 LUTNSDHMTFLQVRKMYPLALSYCHTFSAPIASLSELF---QREVTVDLLSSAS 253
QY 219 A--LETEOLKAIDRYELISIGPLIPSPISFDGNDPSSNKSYYGDLFRKADTYMDWLNS 276
Db 254 VMLFRSDFVKOYPR-----PIMPNMVFFGGINCANG-----KPLSQEFAYINA 297
QY 277 KPSSVYVVSFGSL--RLPKPQMBEIAIGLSDTKSPVLWIRNNEEGDEQAEERKLL 335
Db 298 SGHEGIVVFSLESVMVSEIPEKKAIAADALGKIPQTVLW----- 336
QY 336 SFFDRHGTERLGK-----IVTWCSDLVLTTHKSVGCFVTHCGWNSATIESLACGVPVYCF 389
Db 337 ----RYTGTSPNSLIANTILVWLPQNDLGHCHPTAFITHAGSHGVYESICNGVPMYMM 392
QY 390 PQWFQDGTNAKIEDVWRKSVYRVNVEGGVDRREIKRCVSEVTKSRELRESAMMKGL 449
Db 393 PLFGQDMQNAKME--TKGAGVTNLVLE---MTSEDLNAQRAVINDRKSYKENIMRLSSL 447
QY 450 AKE 452
Db 448 HKD 450

RESULT 15
R26153
ID R26153 standard; Protein; 533 AA.
AC R26153;
DT 27-JAN-1993 (first entry)
DE HUG-Brl;
KW Bilirubin; UDP-glucuronosyltransferase; HUGBrl; HUGBr2;
KW monoglucuronide; diglucuronide.
OS Homo sapiens.
FH Key
FT region 10..20
FT /note= "putative membrane-insertion signal"
FT region 491..507
FT /note= "putative membrane-anchoring peptide"
FT modified_site 102
FT /note= "predicted Asn-linked glycosylation site"
FT modified_site 295
FT /note= "predicted Asn-linked glycosylation site"
FT modified_site 347
FT /note= "predicted Asn-linked glycosylation site"
FT misc_difference 158
FT /note= "feature not labelled in specification"
FT misc_difference 181
FT /note= "feature not labelled in specification"
FT misc_difference 228
FT /note= "feature not labelled in specification"
PN WO9212987-A.
PD 06-AUG-1992.
PF 10-JAN-1992; U00282
PR 10-JAN-1991; US-639453.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Owens IS, Ritter JK;
DR WPI; 92-284593/34.
DR N-PSDB; Q27369.
PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT types I and II
PS Disclosure; Fig 9A-I; 99pp; English.
CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC been isolated. They are referred to as HUGBrl (Q27369) and HUGBr2
CC (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC upon expression individually in COS-1 cells, encode isoforms that
CC catalyse the formation of the two bilirubin monoglucuronides and
CC the diglucuronide. Identical 3' ends (1469 bp in length) to each
CC The cDNAs contain identical 3' ends (1469 bp in length) to each

CC other and to that of the human phenol transferase cDNA, HLUGP1
CC (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC In contrast, they have unique 5' ends.
SQ Sequence 533 AA;

Query Match 6.38; Score 156; DB 1; Length 533;
Best Local Similarity 23.48; Pred. NO. 7.3e-07;
Matches 52; Conservative 35; Mismatches 89; Indels 46; Gaps 6;

QY 238 PLIPSSIFSDGNDPSSNKSYYGDLFRKADETYMDWLNSKPSSVYVVSFGSL--RLPKP 296
Db 267 PIMPNMVFFGGINCILHQN-----PLSQEFAYINAGSGHGVVFSLESVMVSEIPEK 317
QY 297 OMEETAIIGLSDTKSPVLWIRNNEEGDEQAEERKLLSFFDRHGTERLGK-----IV 350
Db 318 KAMAIADALGKIPQTVLW-----RYTGTSPNSLIANTILV 352
QY 351 TWCSDLVLTTHKSVGCFVTHCGWNSATIESLACGVPVYVYCFQWPDQGTNAKIEDVWRSGV 410
Db 353 KWLQNDLLGHCHPTAFITHAGSHGVYESICNGVPMYMMPLFGQDMQNAKME--TKGAGV 411
QY 411 RYRVNVEGGVVDREIKRCVSEVTKSRELRESAMMKGLAKE 452
Db 412 TLNVLE---MTSEDLNAQRAVINDRKSYKENIMRLSSLHKD 449

Search completed: August 1, 2000, 19:50:24
Job time: 21504 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: August 1, 2000, 19:51:33 ; Search time 59.57 Seconds
(without alignments)
123.004 Million cell updates/sec

Title: US-09-147-955-8
Perfect score: 2484
Sequence: 1 MVNKRHILLATFPAGQHNP.....GSSMNLKNFTRIINENAS 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues
Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/pdata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/pdata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/pdata/2/1aa/6.COMB.pep.*
4: /cgn2_6/pdata/2/1aa/PCFUS.COMB.pep.*
5: /cgn2_6/pdata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	631	25.4	471	2	US-08-466-583-2
2	631	25.4	471	4	PCT-US95-07820-2
3	398.5	16.0	471	3	US-09-106-464-2
4	382	15.4	488	2	US-08-797-226-2
5	325.5	13.1	131	2	US-08-466-583-5
6	325.5	13.1	131	4	PCT-US95-07820-5
7	255.5	10.3	131	2	US-08-466-583-6
8	255.5	10.3	131	4	PCT-US95-07820-6
9	205	8.3	63	2	US-08-466-583-8
10	205	8.3	63	4	PCT-US95-07820-8
11	175.5	7.1	534	4	PCT-US92-00282-4
12	170.5	6.9	531	4	PCT-US92-00282-6
13	170	6.8	529	4	PCT-US92-00282-7
14	168	6.8	533	4	PCT-US92-00282-3
15	166	6.7	531	4	PCT-US92-00282-5
16	148.5	6.0	506	5	5180581-2
17	113	4.5	56	2	US-08-466-583-4
18	113	4.5	56	4	PCT-US95-07820-4
19	96.5	3.9	617	1	US-08-361-920-29
20	96.5	3.9	617	1	US-08-479-939-29
21	96.5	3.9	617	1	US-08-483-432-29
22	96.5	3.9	700	2	US-07-862-588B-2
23	96	3.9	906	1	US-08-687-379-4
24	96	3.9	907	1	US-07-718-575-2
25	96	3.9	907	1	US-08-481-206-2
26	96	3.9	907	1	US-08-486-269A-2
27	94.5	3.8	414	2	US-08-750-524-1
28	94.5	3.8	1338	3	US-08-750-141A-3

29	94	3.8	58	2	US-08-466-583-9	Sequence 9, Appli
30	94	3.8	58	4	PCT-US95-07820-9	Sequence 9, Appli
31	92.5	3.7	3052	2	US-08-557-122A-26	Sequence 26, Appli
32	92	3.7	52	2	US-08-466-583-7	Sequence 7, Appli
33	92	3.7	52	4	PCT-US95-07820-7	Sequence 7, Appli
34	91.5	3.7	74	4	PCT-US92-00282-24	Sequence 24, Appli
35	90.5	3.6	1162	2	US-08-599-455B-43	Sequence 43, Appli
36	89.5	3.6	574	2	US-08-307-166-12	Sequence 12, Appli
37	89	3.6	2710	1	US-08-480-604A-6	Sequence 6, Appli
38	89	3.6	2710	2	US-08-405-496A-6	Sequence 6, Appli
39	86.5	3.5	1311	1	US-08-340-011-5	Sequence 5, Appli
40	86	3.5	906	1	US-08-254-573-2	Sequence 2, Appli
41	86	3.5	906	1	US-08-687-379-2	Sequence 2, Appli
42	85.5	3.4	781	1	US-08-373-134D-2	Sequence 2, Appli
43	85.5	3.4	781	2	US-09-114-637-2	Sequence 2, Appli
44	85.5	3.4	828	1	US-08-261-304-2	Sequence 2, Appli
45	85.5	3.4	1479	2	US-08-951-912-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-466-583-2
; Sequence 2, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Randurski, Robert S.
; APPLICANT: Szczygowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-583-2

Query Match 25.4% Score 631; DB 2; Length 471;
Best Local Similarity 33.7%; Pred. No. 2e-56;
Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

```
QY 6 HILLATPAGQHINSLSEFAKRLNTGYVDQVTFPTSYVALRRMRPDTDPSSRIDFVAXX 65
Db 4 HVLVFPFGQGHNPVQFAKRLASKGVA--TTLVTRFTORTADVDAHPAM---VEAIS 58
QY 66 DSYDDGLKGGDKKNYMSMRKRGTKALKDPLIKLINDAAMGSECYNRVSVVYSHLFSWA 125
Db 59 DGHDEG--GFASAAGVAEYLEKQAAAAASLASLVEARASSA--DAFTCVVYDSYEDWV 113
QY 126 AEVAREVDVPSALLMTIEPATVFDVYFYFENG-----YADDIDAGSDIOQLP-----NL 173
Db 114 LPVARMGLPAVPFSTQSCAVSAYVYHFSQGR LAVPPGAAADGSDGAGAAALSEAFGL 173
QY 174 POLSKODLPSFLPSSPARFRTLMKEKFTDLDKEPKAK-VLINTFDALETEQLKADRY- 231
Db 174 PEMERSELPSEFVDHGP--YPTIAMQAIKQFAHAGKDWLFSFELETEVLGLTKYL 231
QY 232 ELISIGLPLSPSSIFSDNDPSSNK-SYGGDLFRKADETYMDLNSKPSSVYVYSGSL 290
Db 232 KARAGICPCVP--LPTAGRTAGANGRITYGANLV-KPEDACTKWLDTKPDORSVAIVSFGSL 288
QY 291 LRLPKQMEIEAIGLSDTKSPVLVWVIRNEEGDEQQAEBEELKLSFFDRHGHTER-LGKI 349
Db 289 ASLGNQAQKEELARGLLAAGKPFMWVR-----ASDEHQVPYLLAEATATGAAMV 338
QY 350 VTWCSDLDVLTHTKSGVCFVTHCGWNSAIESLACGVPVYCFPQWFDQGTNAKMIEDVYRSG 409
Db 339 VPMCPQLDVLHPAVGCFVTHCGWNSLLEALSFGVPVYALMTDQPTNARNVELAWAG 398
QY 410 VYRVNVEGVDVDRREIKRCVSEVKSRE----LRESAMWKKGLAKEMDEERGSMNKL 465
Db 399 VRARRDAGAGVFLRGVEVCRAVMDGGEAASAAKAAAGWRDRARAAY-APGSSDRNL 457
QY 466 KNFI 469
Db 458 DEFV 461

RESULT 2
PCT-US95-07820-2
; Sequence 2, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczylowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
```

```
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07820-2

Query Match 25.4%; Score 631; DB 4; Length 471;
Best Local Similarity 33.7%; Pred. No. 2e-56;
Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps 16;

QY 6 HILLATPAGQHINSLSEFAKRLNTGYVDQVTFPTSYVALRRMRPDTDPSSRIDFVAXX 65
Db 4 HVLVFPFGQGHNPVQFAKRLASKGVA--TTLVTRFTORTADVDAHPAM---VEAIS 58
QY 66 DSYDDGLKGGDKKNYMSMRKRGTKALKDPLIKLINDAAMGSECYNRVSVVYSHLFSWA 125
Db 59 DGHDEG--GFASAAGVAEYLEKQAAAAASLASLVEARASSA--DAFTCVVYDSYEDWV 113
QY 126 AEVAREVDVPSALLMTIEPATVFDVYFYFENG-----YADDIDAGSDIOQLP-----NL 173
Db 114 LPVARMGLPAVPFSTQSCAVSAYVYHFSQGR LAVPPGAAADGSDGAGAAALSEAFGL 173
QY 174 POLSKODLPSFLPSSPARFRTLMKEKFTDLDKEPKAK-VLINTFDALETEQLKADRY- 231
Db 174 PEMERSELPSEFVDHGP--YPTIAMQAIKQFAHAGKDWLFSFELETEVLGLTKYL 231
QY 232 ELISIGLPLSPSSIFSDNDPSSNK-SYGGDLFRKADETYMDLNSKPSSVYVYSGSL 290
Db 232 KARAGICPCVP--LPTAGRTAGANGRITYGANLV-KPEDACTKWLDTKPDORSVAIVSFGSL 288
QY 291 LRLPKQMEIEAIGLSDTKSPVLVWVIRNEEGDEQQAEBEELKLSFFDRHGHTER-LGKI 349
Db 289 ASLGNQAQKEELARGLLAAGKPFMWVR-----ASDEHQVPYLLAEATATGAAMV 338
QY 350 VTWCSDLDVLTHTKSGVCFVTHCGWNSAIESLACGVPVYCFPQWFDQGTNAKMIEDVYRSG 409
Db 339 VPMCPQLDVLHPAVGCFVTHCGWNSLLEALSFGVPVYALMTDQPTNARNVELAWAG 398
QY 410 VYRVNVEGVDVDRREIKRCVSEVKSRE----LRESAMWKKGLAKEMDEERGSMNKL 465
Db 399 VRARRDAGAGVFLRGVEVCRAVMDGGEAASAAKAAAGWRDRARAAY-APGSSDRNL 457
QY 466 KNFI 469
Db 458 DEFV 461

RESULT 3
US-09-106-464-2
; Sequence 2, Application US/09106464
; Patent No. 601145
; GENERAL INFORMATION:
; APPLICANT: Steffens, John C.
; APPLICANT: Chanas, Gurdev S.
; APPLICANT: Kuai, Jian-Ping
; APPLICANT: Eannetta, Nancy
; TITLE OF INVENTION: Chain Length Specific UDP-Glc-Fatty Acid
; TITLE OF INVENTION: Glucosyltransferases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Bads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

22;

```

; REFERENCE/DOCKET NUMBER: 0011.97
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (510) 559-5731
;
; TELEFAX: (510) 559-5777
;
; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 488 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-08-797-226-2

```

Query Match 15.4%; Score 382; DB 2; Length 488;
Best Local Similarity 23.5%; Pred. No. 9e-31;
Matches 127; Conservative 92; Mismatches 186; Indels 136; Gaps 18;

Qy	6	HILLATPAGOHINPSLEFAKELLNTGYVDQVTFSTSVYALRRMRFETDPSSRID-----	60
Db	12	HVLFLPFLSAGHFIPLVNARLFASRGV--KATILTPHNALLFRSTIDDDVRSFGPIS	69
Qy	61	FVAXXDSYDDGLKGDDCKNWS-EMRKR-----GTKALDXTLLKLNDAAMGSECY	110
Db	70	IVTIKPSAEVGLPEGIESFNAGTSPMPHKIFYALSLLQPMEDKIRELPDQIFSDMY	129
Qy	111	NRYSFVYISLFSWAAEVARVDPSALLWIEPATVDFVYFYFNGYADDIDAGSDQIQ	170
Db	130	-----FPWTVDIADELHIPRILYNLSAYWCYSIMH-----NLKVYR	165
Qy	171	PNLPQLSKQDLPSFLPSSPAPRFTLMKEKFTDLQKEPAK-----	211
Db	166	PH-KQPNLDESQVFPVGPIDBIKPKLSQTLDDLRKSDQOKTVDFDELLEQVEDSEERSYG	224
Qy	212	VLINTFDALETEOLKADRYELIS-----IGPLI-----PSSIFSDGNDPSSNNKSY	258
Db	225	IVHDTYELEPAY---VDYQKLKPCKWHFGPLSHFASKIRSKELISEHN-----	273
Qy	259	GGDLFRKADTYNDMLNKPSSVYVYVSGSLLRLLPKPOMBEIATIGLSDTKSPVLWVIRR	318
Db	274	-----NEIVDMLNACKPKSVLYVSGSMARFPESQNLNEIAQALDASNVPTFVLRP	325
Qy	319	NEE-----GDSEQAEAEKLLSFFDRHGHTERLGKIVTWCSOLDVLTHKSYGCFVTHC	371
Db	326	NETASWLPVGNLEDK-----FKGLYIKGWVPQLTIMHEGATGFMFTHC	370
Qy	372	GWNSAIESLACGVPVYVCFQWFQDQGTNAKNIE-----DYMRSQVRVYVNEEGVY	421

RESULT 4
US-08-797-226-2
; Sequence 2, Application US/08797226

Qy	319	NEE-----GDREQAEEEKLLSFFDRHGRTERLKVTVWCSQLDLVTHKSGVCFVTHC	371
Db	326	NETASWLPVGNLDDK-----TKGLYTKGWPPQITIMEHSATGFMFTHC	370
Qy	372	GWNSATSLACGVPVVCFFQWFDQGTNAKME-----DWRSGVRVRVNEEGGVV	421

```
Db 371 GNSVLEAITFCVPMITWPLADFYNEKVVVEVRLGKIGKIDVWNEGIEI-----TGPVI 426
QY 422 DREIKRCVSEVSKR-----ELRESAMWKGGLAKPAMDEERGSSMNNLKNPITRIINE 475
Db 427 ESAKIREATERLMSNGSEETINIRDRVAMSKMAQNATN-EGGSSWNNLTALIQHKNY 485
QY 476 N 476
Db 486 N 486

RESULT 5
US-08-466-583-5
; Sequence 5, Application US/08466583
; Patent No. 5919998
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczylowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,583
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
;
US-08-466-583-5

Query Match 13.1%; Score 325.5; DB 2; Length 131;
Best Local Similarity 46.4%; Pred. No. 6.4e-26;
Matches 64; Conservative 18; Mismatches 45; Indels 11; Gaps 2;

QY 273 WLNSKPSSVVVYVSGSLRLPKPOMEIATGLSDTKSPVLWVIRNEEGDEQEAEER 332
Db 4 WLDTKPDRSVATVYVSGSLASLGNQAQKELARGLAAGKPLFWVVR-----ASDEH 53
QY 333 KLLSPFDRHGTGR-LGKIVTWCSQLDVLTHKSVGCFVTHCGWNSAIESLACGVPVVCPPQ 391
Db 54 QVPRYLLAETATGAAMVVPNCPLDVLHAHPAVGCFVTHCGWNSTLEALSFGVPMYAMAL 113

Db 371 GNSVLEAITFCVPMITWPLADFYNEKVVVEVRLGKIGKIDVWNEGIEI-----TGPVI 426
QY 422 DREIKRCVSEVSKR-----ELRESAMWKGGLAKPAMDEERGSSMNNLKNPITRIINE 475
Db 427 ESAKIREATERLMSNGSEETINIRDRVAMSKMAQNATN-EGGSSWNNLTALIQHKNY 485
QY 476 N 476
Db 486 N 486

RESULT 7
```

```
QY 392 WFDQGTNAKMIEDVWRSG 409
Db 114 WTDQPTNARNVELAWGAG 131

RESULT 6
PCT-US95-07820-5
; Sequence 5, Application PC/TUS9507820
; GENERAL INFORMATION:
; APPLICANT: Bandurski, Robert S.
; APPLICANT: Szeszen, Jędrzej B.
; APPLICANT: Szczylowski, Krzysztof
; TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
; TITLE OF INVENTION: and Plant Growth.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07820
; FILING DATE: 19-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,427
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 11-94B PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
;
PCT-US95-07820-5

Query Match 13.1%; Score 325.5; DB 4; Length 131;
Best Local Similarity 46.4%; Pred. No. 6.4e-26;
Matches 64; Conservative 18; Mismatches 45; Indels 11; Gaps 2;
```

```
QY 273 WLNSKPSSVVVYVSGSLRLPKPOMEIATGLSDTKSPVLWVIRNEEGDEQEAEER 332
Db 4 WLDTKPDRSVATVYVSGSLASLGNQAQKELARGLAAGKPLFWVVR-----ASDEH 53
QY 333 KLLSPFDRHGTGR-LGKIVTWCSQLDVLTHKSVGCFVTHCGWNSAIESLACGVPVVCPPQ 391
Db 54 QVPRYLLAETATGAAMVVPNCPLDVLHAHPAVGCFVTHCGWNSTLEALSFGVPMYAMAL 113

QY 392 WFDQGTNAKMIEDVWRSG 409
Db 114 WTDQPTNARNVELAWGAG 131
```

RESULT 7


```

1  TITLE OF INVENTION: Genetic Control of Plant Hormone Levels
2  TITLE OF INVENTION: and Plant Growth.
3  NUMBER OF SEQUENCES: 9
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Greenlee & Winner, P.C.
6  STREET: 5370 Manhattan Circle, Suite 201
7  CITY: Boulder
8  STATE: Colorado
9  COUNTRY: United States of America
10 ZIP: 80303
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patentin Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: PCT/US95/07820
19 FILING DATE: 19-JUN-1995
20 CLASSIFICATION:
21
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/265,427
24 FILING DATE: 24-JUN-1994
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Ferber, Donna M.
27 REGISTRATION NUMBER: 33,878
28 REFERENCE/DOCKET NUMBER: 11-94B PCT
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 303/499-8080
31 TELEFAX: 303/499-8089
32 TELEX: 49617824
33 INFORMATION FOR SEQ ID NO: 6:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 131 amino acids
36 TYPE: amino acid
37 STRANDEDNESS: single
38 TOPOLOGY: linear
39 MOLECULE TYPE: protein
40 HYPOTHETICAL: NO
41 FRAGMENT TYPE: internal
42
43 PCT-US95-07820-6

```

[illegible]

```

RESULT          9
US-08-466-583-8
: Sequence 8, Application US/08466583
: Patent No. 591998
: GENERAL INFORMATION:
: APPLICANT: Bandurski, Robert S.
: APPLICANT: Szerszeni, Jędrzej B.
: APPLICANT: Szczygłowski, Krzysztof
: TITLE OF INVENTION: Genetic Com
: TITLE OF INVENTION: and Plant G
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee & winner,
: STREET: 5370 Manhattan Circle

```


[illegible]

RESULT 13
PCT-US92-00282-7
Sequence 7, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT

[illegible]

Search completed: August 1, 2000, 19:51:34
Job time: 18756 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:32:19 ; Search time 31.57 Seconds
(without alignments)
469.205 Million cell updates/sec

Title: US-09-147-955-8
Perfect score: 2484
Sequence: 1 MVNKRHILLATFPAGQHNP.....GSSMNLKNFTRIINENAS 478

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	631	25.4	471	1	IAAG_MAIZE
2	453	18.2	453	1	UFOG_GENTR
3	420	16.9	433	1	UFOG_SOLME
4	382	15.4	454	1	ZOX_PHALU
5	373	15.0	459	1	ZOX_PHALU
6	357.5	14.4	449	1	UFO1_MANES
7	354	14.3	487	1	UFO5_MANES
8	349.5	14.1	346	1	UFO2_MANES
9	340	13.7	287	1	UFO7_MANES
10	326.5	13.1	471	1	UFO3_MAIZE
11	325.5	13.1	471	1	UFO1_MAIZE
12	324	13.0	473	1	UFOG_PETHY
13	323.5	13.0	394	1	UFO6_MANES
14	323.5	13.0	471	1	UFO2_MAIZE
15	313.5	12.6	455	1	UFOG_HORVU
16	281.5	11.3	241	1	UFO4_MANES
17	240	9.7	154	1	UFOG_VITVI
18	184.5	7.4	535	1	UD11_RAT
19	182	7.3	535	1	UD11_MOUSE
20	175.5	7.1	534	1	UD14_HUMAN
21	174.5	7.0	533	1	UD12_MOUSE
22	172.5	6.9	532	1	UD14_RABIT
23	170.5	6.9	533	1	UD12_RAT
24	170.5	6.9	534	1	UD15_HUMAN
25	170	6.8	529	1	UD16_RAT
26	170	6.8	531	1	UD16_MOUSE
27	169	6.8	530	1	UD18_RAT
28	167.5	6.7	534	1	UD13_HUMAN
29	167	6.7	533	1	UD11_HUMAN
30	166.5	6.7	531	1	UD16_RABIT
31	166	6.7	531	1	UD16_HUMAN
32	165.5	6.7	520	1	UD17_MOUSE
33	163	6.6	530	1	UD12_HUMAN

34	163	6.6	531	1	UD13_RAT
35	161.5	6.5	531	1	UD15_RAT
36	160.5	6.5	531	1	UGTH_CAEEL
37	160	6.4	530	1	UDBH_MACFA
38	160	6.4	541	1	CGT_HUMAN
39	155	6.2	531	1	UD17_RAT
40	154	6.2	488	1	UDPE_NPVLD
41	154	6.2	541	1	CGT_MOUSE
42	154	6.2	541	1	CGT_RAT
43	153.5	6.2	491	1	UDPE_NPVCF
44	153.5	6.2	529	1	UDB7_HUMAN
45	153	6.2	430	1	OLED_STRAT

ALIGNMENTS

RESULT 1

IAAG_MAIZE					
ID	IAAG_MAIZE	STANDARD;	PRT;	471	AA.
AC	Q41819;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	INDOLE-3-ACETATE BETA-GLUCOSYLTRANSFERASE (EC 2.4.1.121) (IAA-GLU				
DE	SYNTHETASE) ((URIDINE 5'-DIPHOSPHATE-GLUCOSE:INDOL-3-YLACETYL)-BETA-D-				
DE	GLUCOSYL TRANSFERASE).				
GN	IAGLU.				
OS	Zea mays (Maize).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.				
FN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.				
RC	TISSUE-ENDOSPERM;				
RX	MEDLINE; 94367368.				
RA	Szerszen J.B., Szczygowski K., Bandurski R.S.;				
RT	*Iaglu, a gene from Zea mays involved in conjugation of growth				
RT	hormone indole-3-acetic acid.;				
RL	Science 265:1699-1701(1994).				
CC	-1- CATALYTIC ACTIVITY: UDP-GLUCOSE + INDOL-3-ACETATE = UDP + INDOL-3-ACETYL-BETA-1-D-GLUCOSE				
CC	-1- PATHWAY: FIRST STEP IN INDOL-3-ACETIC ACID CONJUGATION PATHWAY.				
CC	-1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL; L34847; AAA59054.1; -				
DR	MAIZEDB; 83603; -				
DR	PFAM; PF00201; UDPGT; 1.				
DR	PROSITE; PS00375; UDPGT; 1.				
KW	Transferase; Glycosyltransferase.				
FT	DOMAIN 80 85 POLY-ALA.				
SQ	SEQUENCE 471 AA; 49710 MW; 99A7F56C2BB4EB39 CRC64;				

Query Match	25.4%;	Score 631;	DB 1;	Length 471;
Best Local Similarity	33.7%;	Pred. No. 1.6e-39;		
Matches 163;	Conservative 78;	Mismatches 197;	Indels 46;	Gaps 16;
QY	6	HILLATFPAGQHINPSLEFAKRLRLNTGYVDQVTFSTVSVALRRMRFPETDPSRIDFVAXX	65	
DB	4	HVLVVPFGQGHMNMVQFAKRLASKGVA--TLVTTRFIQTADVDAPAM--VEAIS	58	
QY	66	DSYDGLKKGDDGKNYKMRKTKTKDITLKLNDAAAGSECVNRYSVFVYSHLFSWA	125	
DB	59	DGHDEG---GFASAAGVAEYLEKQAAASLASLIVEARASSA--DAFTCVYDSDYDQV	113	

```
QY 126 AEVAREVDVPSALLNIEPATVVDVYFYFNG-----YADDIDAGSDQIQLP-----NL 173
Db 114 LPVARRMGILPVPFFSTQCAVSAYVYHFSGRLAVPFGAAAGSDGAGAAALSEAFGL 173
QY 174 POLSKODLPSPFLSPSPARFTLTKERKFDLTKERPAK-VLINTFOALETEQLKADRY- 231
Db 174 PEMERSELSFVFDHGP--YPTIAQAIKQFAHAGKDWLFLNSFELETEVLAKITKL 231
QY 232 ELISIGLPISSIFSDGNDPSSNK-SYGGDLFRKADETTMWLNKSPSSVYVYFGL 290
Db 232 KARATGCPVP--LPTAGRTAGANRITYGANLV-KPEDACTKWLDTKPDORSAYVYFGL 288
QY 291 LRLPKPQMEIEAIGLSDTKSPVLVIRNNEGDQEQAEKEELKLSFFDRHGTGR-LGRI 349
Db 289 ASIGNAQKEELAGLLAAGKPLWVR-----ASDEHQPVRYLLAEATATGAAMV 338
QY 350 VWCSDLVLTHTKSGVCFVTHCGWNSALESACGVPVCPQPFQDGTNAKMTEDVWRSG 409
Db 339 VPMCPOLDVLAHPAYGCVFTVTHCGWNSLEALSFGVPVMAWALMTDQPTNARNVELAWAG 398
QY 410 VVRVNEEGVDRREIKRCYSEVIKRS-----LRESAMMKGLAKEAMDEERGSMNML 465
Db 399 VRARRDAGAGVLRGEVERCVRAYMDGGEAASAARKAAGWRDRARAIV-APGSSDRNL 457
QY 466 KNFI 469
Db 458 DEPV 461

RESULT 2
UFOG_GENTR STANDARD; PRT; 453 AA.
AC Q96493;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE).
OS Gentiana triflora.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Gentianales;
OC Gentianaceae; Gentiana.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PETAL;
RX MEDLINE; 96416435.
RA Tanaka Y., Yonekura K., Fukuchi-Mizutani M., Fukui Y.,
RA Fujiwara H., Ashikari T., Kusumi T.;
RT Molecular and biochemical characterization of three anthocyanin
RT synthetic enzymes from Gentiana triflora.;
RL Plant Cell Physiol. 37:711-716(1996).
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D85186; BAA12737.1; -
CC FRAM; PF00201; UDPGT; 1.
CC PROSITE; PS00375; UDPGT; 1.
CC Transferase; Glycosyltransferase.
CC SEQUENCE 453 AA; 50009 MW; BF738B0A2DA76C05 CRC64;
```

```
Query Match 18.2%; Score 453; DB 1; Length 453;
Best Local Similarity 26.6%; Pred. No. 2.4e-26;
Matches 137; Conservative 90; Mismatches 162; Indels 126; Gaps 19;

QY 6 HILLATFPAGHINSLSEFAKRLNTGYVDVTFE-----TSVYALRRM----- 49
Db 6 HVALAPPFGTAAAPLLTLVNLAAASADPIIFSFSSSSSTITTFSTNLSIGSNIKPY 65
QY 50 -----REFETDPSRRIDFV--AXXDSYDDGLKKKGDDGKNYSEMRKRGTKALKDYL 97
Db 66 AVWDGSPGEGFVSGNPREPIEYFLNAAPDNFDKAMK-----KAVEDTG 108
QY 98 IKLANDAMSECYNRSFVYVSHLFSW-AAEVAREVDVPSALLNIEPATVVDVYFYFNG 156
Db 109 VNTS-----CLTDAFLWFAADFSEKIGVPWIPVWTAASCSLCUHV----- 149
QY 157 YADDIDAGSDQIQ-----LPNLPOLSKQDLP-SFLFPSSPARFETLTKERKFDL 204
Db 150 YTDEIRKREAFEDIAEKAKTIDFIPGLSAISFSOLDPEELIMEDSQSIFALTLLHNMGLK 209
QY 205 DKEPKAKVLINTFDAL-----TEQLKADRYELISIGPL--IPSSIFSDGNDPSSNKS 259
Db 210 HK--ATAVAVNSFEIDPIITNHLRSTNQLNLMIGPLQTLSSIPPEDNE----- 258
QY 260 GDLFRKADETYMDNLNPKSPSSVYVYVSGSLRLRLPKPQMEIEAIGLSDTKSPVLVIRN 319
Db 259 -----CLKWLQTOKESSVYVLSFGTVINPPNEMAALASTLSRKIPFLWSLR-- 306
QY 320 EEGDEQEQAEKEELK--SFFDRHGTGRILKIVTWCSDLVLTHTKSGVCFVTHCGWNSAI 377
Db 307 -----DEARKHLPEFIDR--TSTFGKIVSWAQHVLNPAIGVFTVTHCGWNSLT 355
QY 378 ESLACGVPVYVCPQPFQDGTNAKMTEDVYRSGVVRVNEEGVDRREIKRCYSEVI--- 434
Db 356 ESIFCRVPVIGRPFQDQKVNARMVEDVWKIVGVK---GGVFTEDETTRVLELVLESD 411
QY 435 KSRELRESAMMKGLAKEAMDEERGSMNMLKNFI 469
Db 412 KGRMQNVRGLKEKAKDAV-KANGSSTRNFESLL 445

RESULT 3
UFOG_SOLME STANDARD; PRT; 433 AA.
AC Q43641;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE).
GN GT OR UGT76.
OS Solanum melongena (Eggplant) (Aubergine).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
OC Solanaceae; Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SINSAOHARANASU; TISSUE=HYPOCOTYL;
RC Toguri T.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```


RA MEDLINE; 99292951.
RA Martin R.C., Mok M.C., Mok D.W.S.;

DT 15-FEB
DT 15-FEB

DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)


```

QY 169 QLPNLPOLSKODLPSPFLPSSPARFRTLMKEKFDYLDKEPK-----AK-VLINTFDALETE 223
Db 159 QVPGI-----VNSPFSKAMPTA-----ILSKQWPPPLLENTTRYGAKGVIINTFELESH 209
QY 224 QLKADRYELISGIPSPISFSDGNDPSSSNKSYGGDLFRKADETYMDLNSKPSSSVV 283
Db 210 AIESFKDPIPVGPIL--DVRSNG-----RNTQNEIQWLDDQPPSSVV 252
QY 284 YVSGSLRLPKPOMEBIAIGLSDTKSPVLWIRNEEGDEOEOQAEKEKL-----SFF 338
Db 253 FLCSGNSFSKQVKEIAACALSDSGHRLWLSLADHRAPGLESPSYDEIQLEVPBGFL 312
QY 339 DRHETERLGRIVTWCSDLDVLTHTKSGVCFVTHCGWNSAIESLACQVPVCPQWFDGDN 398
Db 313 ER--TSGIEKVICWAPVAVLAHPATGGLVSHSGWNSILESIFGVPVATWPMYAEQFN 370
QY 399 AKHIEDVWRSQVRV-----NEEGVVDREIKRCVSEVYKSRRELRESAMWGLAKEA 453
Db 371 A--FQWIELGLAVEIKMDYRDSGEIVKCDQIERGIRCLAKHDSDRKKV-----KEM 422
QY 454 MDEERGSMN-----LKNFIRII 473
Db 423 SEKSGALMEGSSYCHLDNLKDMI 448

RESULT 7
UFO5_MANES STANDARD; PRT; 487 AA.
AC Q40287;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 5 (EC 2.4.1.91) (UDP-GLUCOSE
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 5).
GN GT5 OR UGT73A5.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
OC Euphorbiaceae; Manihot.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-COTYLEDON;
RX MEDLINE; 95201291.
RA Hughes J., Hughes M.A.;
RT "Multiple secondary plant product UDP-glucose glucosyltransferase
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.";
RL DNA Seq. 5:41-49(1994).
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -!- TISSUE SPECIFICITY: FAINTLY EXPRESSED IN COTYLEDONS.
CC -!- DEVELOPMENTAL STAGE: RARE TRANSCRIPTS EXPRESSED IN COTYLEDON AND
CC ROOTS DURING THE DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
DB EMBL; X77462; CAA54612.1; -.
DR PFAM; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Multigene family.
SQ SEQUENCE 487 AA; 54381 MW; 0F5CA2EBC897F124 CRC64;

```

```

Query Match 14.38; Score 354; DB 1; Length 487;
Best Local Similarity 24.68; Pred. No. 5.7e-19;
Matches 134; Conservative 88; Mismatches 171; Indels 152; Gaps 22;

QY 3 NKRHTLLATPPAOGHINPSLEFAKRLNTGTYVDVTF-----TSVYALRRMRRETD-- 55
Db 8 SKPHVLLSSPGLGHLIPVLEKRIIVTLCNFD-VTFWVGSDTSAAPQVURSANTPKL 66
QY 56 -----SSRID-----FVAXXSDYDGLKKGDDGKNYMSMKRGTALKUD 95
Db 67 CEIIQLPPNISCULDPEATVCTRLFV-----LMREIRPAFAAAV-- 106
QY 96 TLILNDAAMSGECYNRVSVVYSHLFSNAAEVAREVDVP-----SALLWIEPAVFDVY 150
Db 107 SALKRPAIIIVDLFGTESL-----EVAKEGLAKVYVYASNAWFLALTYI--- 152
QY 151 YFPNGYADDIDAGSDQIQLPNLPOLSKODLPSPFLPSS-----PARFRTLMKEKFD 202
Db 153 -----VPILDKEVEGEFVLQKEPKMIPGCRPVRTVEVVDPMLD 190
QY 203 TLDKE-----PKAK-VLINTFDAL-----TEQLKAIDRYELISGIPLP 241
Db 191 RTNQYSEYFRLGIEIPTADGILMTWEALEPTFGALRDVFLGRVAKVPVPIGDLRR 250
QY 242 SSIFSDGNDPSSSNKSYGGDLFRKADETYMDLNSKPSSVYVYSGSLRLPKPQMBEI 301
Db 251 QA-----GPCGSNCE-----LLDLDQOPKESVYVYSGSGTLLSEQMTEL 292
QY 302 ATGLSDTKSPVLWIRN--EEGDEQ--EQAEERELLSFFDRHGHTERL--GKTV-TWC 353
Db 293 ANGLERSOQRIWVWVROPTVTKTGDAFFTGCGADMSGYFEGFLTRIONVGLVVPWS 352
QY 354 SOLDLVTHKSGVCFVTHCGWNSAIESLACQVPVCPQWFDGDNKMHIEDVWRSQVRV 413
Db 353 PQIHIMSPSGVFLSHCGWNSVLESTAGVPIIAWPIYAEQRMNATLLTEELGVAVRPK 412
QY 414 VNEEGVVDREIKRCVSEVI---KSRLESAMWGLAKEAMDEERGSSMNNLKNTIT 470
Db 413 NLPKVVKREIERMIRIMVDEEGSEIKRVRELKDSGEKALN-EGSSSF----NYMS 467
QY 471 RIINE 475
Db 468 ALGNE 472

```

```

RESULT 8
UFO2_MANES STANDARD; PRT; 346 AA.
AC Q40285;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 2 (EC 2.4.1.91) (UDP-GLUCOSE
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 2) (FRAGMENT).
GN GT2 OR UGT73A2.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
OC Euphorbiaceae; Manihot.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-COTYLEDON;
RX MEDLINE; 95201291.
RA Hughes J., Hughes M.A.;
RT "Multiple secondary plant product UDP-glucose glucosyltransferase
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.";
RL DNA Seq. 5:41-49(1994).
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL

```


Db 13 HVAVAFPFSSHAALLSIARALAAAPSGATLSFLSTASSLAQLRKASSAGHGLPG 72
QY 58 RIDFVAXXSDYDGLKDDG-----KNYSEMRRRGYKALKDTLIKUNDAAGSECY 110
Db 73 NLRV-----EVPDGAPEAEVVPVPMOLFMEEAAEAGGKA-----WLEAARAAAGGA-- 122
QY 111 NRVSEFVYSHLFSW-AAEAREVDVPSALLWIEPATVDFVYFVNGYADDI-DAGSDO- 167
Db 123 -RVCVV-GDAFWPAADAASAGAPWVWTAASCAL-LAHIRTDALREDVQDQANRV 179
QY 168 ----TQPNLPOLSKQDLPSFLLPSSPARFRTLMKEKFDLDPKPKAKVLIINTFDATELE 223
Db 180 DGLLISHPLGASRYRDLPGVYSGDFYVNLVHVMGQCLPRSAVAALNTPFGLDPP 239
QY 224 QLKAIIDRYELISIGPLIPSSI-----FSDGNDPSSSKSVGGDLPRKADETMYDL 274
Db 240 DVTA-----ALAEILPNCVPGFPHLLAEDDADTAAPADPHG-----CLAWL 282
QY 275 NSKPESVYVYSGSLRLPKP-QMEETAIGLSDTKSPVLWVIRNNEEGDEQDAEERK 333
Db 283 GROPARGVAVSFGT-VACPRDELRELAAGLSDSGAPFLWSLR-----EDSWP 330
QY 334 LL--SEFDRHGTERRGKVTWCQSOLDVLTHKSGVCFYTHCGWNSAIESLACGVPVVCFPQ 391
Db 331 HLPFGFLDRAAGTSGLVVPAPQVAVLHPSVGAFTHAGWASVLEGLSSGVPMACRPF 390
QY 392 WFDQGTNAKMIEDVNRGVRVVRNVEEGVDREIKRCVSEVSKSRE---LRESANMMKG 448
Db 391 FGDQRMNARSVAHVNGFAAFE-----GAMTSAGVATVABELLRGEBGARMARAKELQA 445
QY 449 LAKEAMDEERGSSMNLKNFI 469
Db 446 LVAREFG-PGGECKRNFDEFV 465

RESULT 12

UFOG_PETHY
ID UFOG_PETHY STANDARD; PRT; 473 AA.
AC Q43716; Q40895;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
3-O-GLUCOSYLTRANSFERASE) (ANTHOCYANIN RHAMNOSYL TRANSFERASE).
GN RT OR UGT72.
OS Petunia hybrida (petunia).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
OC Solanaceae; Petunia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. V-26; TISSUE=LEAF;
RX MEDLINE; 94177177.
RA Kroon J., Souer E., de Graaff A., Xue Y., Mol J., Koes R.;
RT "Cloning and structural analysis of the anthocyanin pigmentation
locus Rt of Petunia hybrida: characterization of insertion sequences
in two mutant alleles";
RL Plant J. 5:69-80(1994).
RN [2]
RP SEQUENCE OF 3-473 FROM N.A.
RC STRAIN=CV. OLD GLORY BLUE; TISSUE=PETAL;
RX MEDLINE; 94177178.
RA Brugliera F., Holton T.A., Stevenson T.W., Farcy E., Lu C.Y.,
RA Cornish E.C.;
RT "Isolation and characterization of a cDNA clone corresponding to the
Rt locus of Petunia hybrida";
RL Plant J. 5:81-92(1994).
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
PIGMENTS.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE

CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71060; CAA50377.1; -;
DR EMBL; X71059; CAA50376.1; -;
DR EMBL; X25802; CAA81057.1; -;
DR PFAM; PF00201; UDPGT; 1;
DR PROSITE; PS00375; UDPGT; 1;
KW Transferase; Glycosyltransferase.
FT CONFLICT 246 255 LIGPVVDPDP -> SNRTRSGPA (IN REF. 1).
FT CONFLICT 348 348 N -> H (IN REF. 2).
FT CONFLICT 431 431 D -> E (IN REF. 2).
SQ SEQUENCE 473 AA; 52405 MW; A310E4BFC9EB02A CRC64;
Query Match 13.0%; Score 324; DB 1; Length 473;
Best Local Similarity 24.3%; Pred. No. 9.1e-17;
Matches 120; Conservative 82; Mismatches 217; Indels 74; Gaps 16;
QY 6 HILLATPAGQHINPSLEPAKRLNTGYVDQVTFSTVYALRRMRFTDPSSRIDFVAXX 65
Db 13 HVMPFFFAFHISPPVQLANKL--SSYGVKVSFFTASGNASRVSMLNSAPTTHIVPLT 70
QY 66 DSYDDGLKKGDKGNYSMMRKRTKALKDTLIKUNDAAGSECVNVSFVYSHLPSWA 125
Db 71 LPHEGLUPGAES---TAEITPASAEILKVALDLMQPOIKTLLSHLKPHEVDFDFAQEWL 127
QY 126 AEVAREVDVPSALLWIEPATVDFVYFVNGYADD-IDAGSDQIQLPNLQPSKQDLPSF 184
Db 128 PKMANGLGITK-----VYYSVVALSTAFITCPARVLEPKKYPSEDMKKPPL 175
QY 185 LLP-SSPARFRTLMKEF-----DTLDKEPK--AKVLINTDALETOLKA 227
Db 176 GFQTSVTSVTFEARDFLYVFKSFHNGFTLYDIQSLRCGSAI LAKTCSQEGPIKY 235
QY 228 ID---RYELISIGLPIPSIFSDGNDPSSSKSVGGDLFRKADETYMDLNSKPESVY 284
Db 236 VEAQFNKPFVIGPVVP-----DPPSG-----KLEEKWATLKNKEGGTVIY 277
QY 285 VSGSLRLPKPOMEBIAIGLSDTKSPVLWVIRNNEEGDEQDAEERKLLSPFDRHGTE 344
Db 278 CSFGSETFTDDQVKELALGLEQTGLPFELV--NFPANVDVSAELNALPEGF---LE 331
QY 345 RL---GKIYV-WCSQDLVLTHKSGVCFYTHCGWNSAIESLACGVPVVCFFQWDFDQGTNAK 400
Db 332 RVKDKGIHSHGWVQQONILAHSSGVYCHAGFSVIEALVNDQCVVMLPKQGDILNAK 391
QY 401 MIEDVNRSGVRVVRNVEEGVDREIKRCVSEVSKSRE---LRESANMMKG--LAKE 452
Db 392 LVSDMEAGVIEINRDEQYFKEDIKEAVEKVVDVEKDPGKLIRENQKKWEFLNKD 451
QY 453 AMDEERGSSMNL 465
Db 452 IQSKYIGNLVNM 464
RESULT 13
UFOG_MANES
ID UFOG_MANES STANDARD; PRT; 394 AA.
AC Q40288;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 6 (EC 2.4.1.91) (UDP-GLUCOSE
FLAVONOID 3-O-GLUCOSYLTRANSFERASE 6) (FRAGMENT).

GN GT6 OR UGT7346
 OS Manihot esculenta (Cassava) (Manioc).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
 OC Euphorbiaceae; Manihot.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COTYLEDON;
 RX MEDLINE: 95201291.
 RA Hughes J., Hughes M.A.;
 RT "Multiple secondary plant product UDP-glucose glucosyltransferase
 genes expressed in cassava (Manihot esculenta Crantz) cotyledons.";
 RL DNA Seq. 5:41-49(1994).
 CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
 CC PIGMENTS (BY SIMILARITY). MAY BE INVOLVED IN GLYCOSYLATION OF
 CC INSTABLE CYANOHYDRINS TO PRODUCE STABLE CYANOGLUCOSIDES.
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
 CC 3-O-D-GLUCOSIDE.
 CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
 CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS AND LEAVES.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN DEVELOPING COTYLEDONS
 CC (FROM EMERGING GREEN COTYLEDONS TO APPROXIMATELY 10 DAYS OLD);
 CC VERY LOW LEVELS IN HYPOCOTYLS AND NO MEASURABLE EXPRESSION IN
 CC ROOTS.
 CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X77463; CAA54613.1; .
 DR PFAM: PF00201; UDPGT. 1.
 DR PROSITE: PS00375; UDPGT. 1.
 KW Transferase; Glycosyltransferase; Multigene family.
 FT NON_TER 1
 SQ SEQUENCE 394 AA; 44264 MW; FD51CD974E5ADC7D CRC64;

Query Match 13.0%; Score 323.5; DB 1; Length 394;
 Best Local Similarity 27.0%; Pred. No. 7.7e-17;
 Matches 103; Conservative 81; Mismatches 128; Indels 69; Gaps 16;

QY 127 EVAREVDVPSALLWIEPATVFDVYFYFNGYADDIDA-----GSD-QIQLNPLQLSKQ 179
 DB 45 DVAKEGVPIYIFFTSGA-AELGLFTVQLIHDEQDADLTQFKDSDAELSVPSL-----AN 99
 QY 180 DLPSFLPSSPARFRTLMKEFTLD-----REPKAKVLINTFDALTEQLKAI--DRY 231
 DB 100 SLPARVLPAS-----MLVKDRFYAFIRIIRLEAKG-INVTFMELESHALNSLKDDQS 153
 QY 232 E---LISGTLIPSSIFSDGNDPSSNKSXGGDLFRKADTYMDNLNSPESVYVVFSG 288
 DB 154 KIPPYPVPILKSNQNDVCPGSE-----IIEWLDDQPPSSVVFVLCFG 199
 QY 289 SLRLRPKQMEIEAIGLSDTKSPYLWVIRNEEGDEQEQAEEBKL-----LSFFDRHCT 343
 DB 200 SMGGFDMDQAKEACALBQSRHRLWSLRPPPKGIETSTDYENLQETILPVGFSE--T 257
 QY 344 ERIGKLVTCWSOLDVTHKSGCVFTWCGWNSALESACGVPVPCFQWFDQGTNA-KMI 402
 DB 258 AGMKGVGNAPQVALEHPAIGGGVFSVHCGWNSILESIMFVSPATWPLAEQGFNAFTW 317
 QY 403 EDVWRSGVRVRY-----NEEGGVDRRIEIKR--CYSEVIKSRRELRESAMMKGLAKEAM 454
 DB 318 TEL--GLAVEIKMDYKKESEILSADDIERGICKVME--HHSEIRKRYKEMSKSRAL 372
 QY 455 DREGRSSHNLNKFTIRLINE 475

DB 373 MODESSSF-----WLDRLIED 388
 RESULT 14
 UPO2_MAIZE
 ID UPO2_MAIZE STANDARD: PRT; 471 AA.
 AC P16155;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
 DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).
 GN BZ1 OR UGT71A1.
 OS Zea mays (Maize)
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Furtak D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr.;
 RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
 RL Plant Mol. Biol. 11:473-481(1988).
 CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
 CC PIGMENTS.
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
 CC 3-O-D-GLUCOSIDE.
 CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
 CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
 CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X13501; CAA31856.1; .
 DR PIR: S08325; S08325.
 DR MAI2EDB; 13885;
 DR PFAM: PF00201; UDPGT. 2.
 DR PROSITE: PS00375; UDPGT. 1.
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 471 AA; 48621 MW; 81B897410A361299 CRC64;

Query Match 13.0%; Score 323.5; DB 1; Length 471;
 Best Local Similarity 24.2%; Pred. No. 9.9e-17;
 Matches 121; Conservative 76; Mismatches 219; Indels 85; Gaps 21;

QY 6 HILLATFPAGGHINSLEPAKRLNLTGYVDQVT--FTSVYALRRMFETDPS-----57
 DB 13 HVAVVAFPESSHAALLSTARALAAAAAPSGATLSFLSTASSLAQLRKASSASAGHCLPG 72
 QY 58 RIDFVAXXDSYDDGLKGGDG-----KNTWSEMRKRGTKALKDTLIKINDAMSECY 110
 DB 73 NLRFV-----EVPDGAFAAEETVPVPRQMLFMEAAEAGGVKA-----WLEAARAAGA-- 122
 QY 111 NRVSVFYVSHLFSW-AAEVAREVDVPSALLWIEPATVFDVYFYFNGYADDI-DAGSDQ- 167
 DB 123 -RVTGVV-GDAFWPADAASAAGAPVWPVTAASCAL-LAHIRTDLSRLDVGDOANRV 179
 QY 168 -----IQLNPLQSLQDLPSFLPSSPARFRTLMKEFTLDLKEPKAKVLINTFDALTE 223
 DB 180 DEPLISHPGGLASRYRDLDPDGVVSGDFNVISLLVHRMGQCLPRSAVAALNTFPGDPP 239
 QY 224 QLKADIRVELISIGLIPSSI-----FSDGNDPSSNKSXGGDLFRKADETYMDWL 274
 DB 240 DVTA-----ALAEILPNCVPFPGYHLLAEDDADTAAPADPHG-----CLAWL 282
 QY 275 NSKPESVYVFGSLRLRPKP-QMEIEAIGLSDTKSPYLWVIRNEEGDEQEQAEEBKEK 333

```
Db 283 GROPARGVAYVSGT-VACPRDEURLELAAGLEASAPFFLWSLR-----EDSWT 330
Qy 334 LL--SFFDRHGRTERLGIKIVTWCSDLVITHKSVGCFYTHCGWNSAIESLACGVPVVCFFQ 391
Db 331 LLPFGFLDRAAGTSGSLVVPAPQAVLRHPSVGAFTVTHAGWASVLEGVSGVPMACRPF 390
Qy 392 WFDQGTNAKIEWDRSGVRVVRNEEGVDRREIKRCVSEVIKSR--LRESAMWKG 448
Db 391 FGDRNNAKSAHVHVGFGAFT-----GAMTSAGVAAAVELLURGEAGGACMRARAKELQA 445
Qy 449 LAKEANDEERGSSMNNLNFI 469
Db 446 LVAEAFG-PGGECKRNDFEV 465

RESULT 15
UFQG_HORVU STANDARD; PRT; 455 AA.
AC P14726;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1).
GN B21 OR UGT71A2.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CV. AB2331;
RX MEDLINE: 91329682;
RA Wise R.P., Rohde W., Salamini F.;
RT "Nucleotide sequence of the Bronze-1 homologous gene from Hordeum
RT vulgare.";
RL Plant Mol. Biol. 14:277-279(1990).
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15694; CAA33729.1; -
CC PIR; S14919; XUBHFG.
CC PFAM; PF00201; UDPGT. 2.
CC PROSITE; PS00375; UDPGT. 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 455 AA; 47079 MW; FFD8BC2F2103AC9C CRC64;

Query Match 12.6%; Score 313.5; DB 1; Length 455;
Best Local Similarity 24.2%; Pred. No. 5.2e-16;
Matches 119; Conservative 76; Mismatches 220; Indels 77; Gaps 18;

Qy 6 HILLATFPAQGHINSLFAKELLNTGV-DQVTFFTSVYALRMRRETFDPSRIDFVAX 64
Db 7 HIATAVFPSSHAANVLFSAFALAAAAPAGTSLSLFTADNAQAQRLKAGALPGLNRFV-- 64
Qy 65 XDSYDGLKKGGDQ-----KNYMSERKRGTKALKDTLIKINDAMGSECYNRVSV 116
Db 65 --EVPDGVPPGETSCLSPRRMDLFMAAAEAGGVR-----VGL-EAACASAGGARVSCV 115
```

```
Qy 117 VYSHLFSWAAEVAAREVDVPSALLWTEPATVF-----DYVYFYFNGYADDIDAGS 165
Db 116 V-GDAFVMTADAASAAGAPWAVWTAASCALLAHLRTDALRRDV-----GDQAASRA 166
Qy 166 DQIQLPN--LPQLSKQDLPFLPSSPARFRTLMKEKFDTLDKPKKA--KVLINTEDALE 221
Db 167 DELIVAHAGLGGRVVRDLDGVVSGDFNVVLSLVHR--QAORLPKAAATAVALNTPGLD 224
Qy 222 TEQLKAIDRYELISIGLIPSSIFSDGNDPSSNNKSYGGDLFRKADETYMDWLNSKPSS 281
Db 225 PPDIIAALAAELFNCLPLGPHLL-PGAETADTNEAPAD-----PHGCLAWLDRRPARS 278
Qy 282 VVYVSGSLRLLPKPOMERIAIGLSDTKSPVLWVIRRNEEGDEGEQAEKEELLSFFDRH 341
Db 279 VAYVSGTNAATARPDELQELAGLEASGAPFLWSLR-----GVVAAAPRG 323
Qy 342 GTERL-GKIVTWCSDLVITHKSVGCFYTHCGWNSAIESLACGVPVVCFFQWFDQGTNAK 400
Db 324 FLERAPGLVVPWAPQGVLRHAAVGAFTVTHAGWASVMEGVSGVPMACRPFPGDQTMNAR 383
Qy 401 MIEDVVRSGVRVVRNEEGVDRREIKRCVSEVIKSR--LRESAMWKGLAKEAMDEE 457
Db 384 SVASVWGFCGTAF-----DGPMTRGAVANAVATLLRGEDEGERMRKAKQELQAVYKAFEPD 438
Qy 458 RGSSMNNLNFI 469
Db 439 -GGCRKNFDEFV 449
```

Search completed: August 1, 2000, 21:32:20
Job time: 6040 sec

Db	112	KEPVTCLINNAFVPVWCDAEELHIPSAVLWVQSCACLTATAYYYYHHRVLVKFPPTKTEPDI -	170
Qy	163	AGSDIOQLPNLPQLSPKODLPFLPLSSP-ARPPRLMKKEFDLTDKPEKAKVLINIFDALE	221
Db	171	----SVEIPCLPLKHDEISFLPSPSYTAFGDIIDQLKRFENHKSFYLIIDTFPRELE	226
Qy	222	---TEOLKAIRDYELIS-IGPLIPSSIFSOGNDPSSNKSYGGLDFRKADETYMDWLNSK	277
Db	227	KDIMHMSQLCPQATISVPGL-----FKMAQTLSSDVK--GDISEPASC-MEWLDSR	277
Qy	278	PSSVVVYSPGSLRLPKPOMEETATIGLSDTKSPVLWIRNNEEGDEGEQAEKEELLSF	337
Db	278	EPSSVYISFGTIANLQEQEETARGVLSGSLVLWVVRPMEG-----TF	334
Qy	338	FORH----GTERLAKIVTWSQDLVLTHTKSVGFTVHCWNSAIBSLACGVPVVCPPQWF	393
Db	325	VEPHVLPRELEKGIKIVCMCPQOERVLAAHPAIACTFLSHCGWNSMTEALTAGVPVVCPPQWG	384
Qy	394	DGHTNAKMIEDWRSQGVVY-RVNEEGGVVDRIEIKFCVSEVI---KSRELRESANMKGL	449
Db	385	DQVTDVAVLADYFKTGVLRGAAEMIVSRVYAEKLEATVGEKAYELNARRWKAE	444
Qy	450	AKEAMDDEERGSGMNNLKNFTIRINENAS	478
Db	445	A-EAAVADGGSDMNFKEFVDKLVTHVT	472
RESULT	8		
A54739			
indole-3-acetate		beta-glucosyltransferase (EC 2.4.1.121) - maize	
C.Scofield		zea mays (maize)	

C;Accession: A54739
R;Szerzen, J.B.: Szczygłowski, K.; Bandurski, R.S.
Science 265, 1699-1701, 1994
A>Title: laglu, a gene from Zea mays involved in conjugation of growth hor
A;Reference number: A54739; MUID:94367368
A;Accession: A54739
A>Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-471 <SZ>
C;Cross-references: GB:I34847; NID:g548194; PIDN:AAS9054.1; PID:g548195
C;Superfamily: flavonol O3-glucosyltransferase
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match	25.4%	Score 631;	DB 2;	Length 471;
Best Local Similarity	33.7%;	Pred. No. 8.7e-40;		
Matches 163;	Conservative 78;	Mismatches 197;	Indels 46;	Gaps

QY 6 HILLATFEAQGHINPLEFAKRLINTGYDVDTFTFSVVALRMREFDPSSEIDFVAXX 65
 |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 4 HVLVVPFCQHNPWFQFAKRASKGVA--TTLVTREIQTADVADHPAN---VEAIS 58

QY 66 DSYDDGLKGDDGNMYNSEMKRRGTALKDKTLIKLNDAAAMGSECYNRVSVFYSHLFWSA 125
 |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 59 DGHDEG--GFASAAGVAYLEKOAASLASLYEARASSA--DAFTCVVYSYEDMV 113

[illegible]

```

C:Accession: A54739
R:Szerszen, J.B.: Szczygłowski, K.; Bandurski, R.S.
Science 265, 1699-1701, 1994
A:title: laglu, a gene from Zea mays involved in conjugation of growth hor
A:Reference number: A54739; MUID:94367368
A:Accession: A54739
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-471 <SIZE>
A:Cross-references: GB:J34847; NID:g548194; PIDN:AAAS9054.1; PID:g548195
C:Superfamily: flavonol O3-glucosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match      25 48; Score 631; DB 2; Length 471;
Best Local Similarity 33.7%; Pred. No. 8.7e-40;
Matches 163; Conservative 78; Mismatches 197; Indels 46; Gaps

QY 6 HILLATFPAQGHINPLEFAKRLINTGYDVQVTFEYVYALRRMFEFDPSSRIDFVAXX 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 HVLVFPFGQGHMFWQFAKRLASKGVA--ITLVTRTQRTADVDAHPAM---VEAIS 58

QY 66 DSYDDGLKGGDGKNYSEMKRGTCTALKRDTLLKLNDAAMGSECYNRVSVVYSHLPSWA 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 DGHDEG---GFASAGVAYEYLEKQMAAASLASLVEARASSA---DAFTCVYVDYEDWV 113

QY 126 AEVAREVDVPSALLWIEPATVFDVYFYFNG-----YADDIDAGSDQQLP----NL 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 LPVARRMGLPVPVSTQCAVSAYVYHFSOGLAVPPGNAADGSDGAGAAALSEAPGL 173

QY 174 POLSKODLPSEFLPSSPARPRTLMKKEFTDLKPEKAK-VLINTFDALETQKAI DRY - 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 PEMERSELPSEVFDHGP--YPTAMQAIKQFAHAGKDDWLVFNFSPEELETVLAGLT KYL 231

QY 232 ELISITGLIPSSIFSDGNDPSSNK--SYGGDLFRADETYMDLNSKPESVYVYVSGSL 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 KARATGCPV--LPTAGRTAGACNRTYYCANLY-RPEDACTKWLDTFPDRSAVYVSGSL 298

QY 291 LRLPKPQMEETAIGLSDTKSPVLVWVIRNNEEGDEOEQAEEBKLLSFDFDRHGTGR-LGKI 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 ASLGNAAQKEELARGLLAAGKPFLLWYR-----ASDEHQVPYRLLEATATGAAYV 338

```

QY 350 VTWCSOLDVLTHTKSCVFTVHCWNSAISLACGVPVVCFFQWFDQGTNAKMIEDVWRS 409
Db 339 VVMCPQLDVLVLAHPVAGCFVTHCGWNSLTALSGFVPMVANAALWTDQPTNARNVELANGAG 398
QY 410 VVRVNRVEGGVVRREIKRVCVSVIKRS-----LRESAMMKWGLAKEAMDEERGSSMNL 465
Db 399 VRARRDAGAGVLRGEVERCVRAVMDGGEAASAKKAGWDRARAAY-APGGSSDRNL 457
QY 466 KNFI 469
Db 458 DEFV 461
RESULT 9
T00639
hypothetical protein F316.2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00639
R:Federpspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,
; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z14197
A:Accession: T00639
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-460 <FED>
A:Cross-references: EMBL:AC002396; NID:92749918; PID:g2829862; GSPDB:GM00059; ATSP:F316.
C:Genetics:
A:Gene: ATSP:F316.2
A:Map position: 1
A:Introns: 219/3
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 24.9%; Score 618.5; DB 2; Length 460;
Best Local Similarity 33.1%; Pred. No. 7.3e-39;
Matches 161; Conservative 82; Mismatches 184; Indels 59; Gaps 18;
QY 4 KRHLLATFPAQGHINPSLEFAKRLNTGVDQVTFEVSVALRRMRFETDPSRIDFVA 63
Db 9 KHVVLIPYPVQGHLPWQFARLVSQWV--KVITATTTTAYASSI---TTPSLSVPEIS 63
QY 64 XDSYD---DGLKGGDKGNKSEMRRKTKALKDPTLIKLNDAAMGSECYNRVSVVYSH 120
Db 64 --DGFDFPIGI--PGFSVDVTSSEFSLNGSETLTLLIEKFTSDPIDC-----LYIDS 114
QY 121 LFSNAAEVAREVDVPSALLWTEPATVDVYFYFNG--YADDIDAGSDQIOLPNLPOLSKQ 179
Db 115 FLPHGLEVARSHLSAASFNTNLTVCVLRKFSNNGDFPLPADPNSAFPRINGULPSUSD 174
QY 180 DLPSFL---LPSSPARFRTLMK-----EKFTLDKPKAKVLINTFDALETEQ---LKA 228
Db 175 ELPFSVGRHWLTHPEHGRVLLNGFPNHNADWL-----FVNGFEGLEETQDCENGES 226
QY 229 DRYELISIGLPLIPSSIFSDGNDSSSKSYGGDLPRKADETYMWLNKSPSSVYVYSG 288
Db 227 DAMKATLIGPMIPSAYL---DDRMEDDKDTGASLLKPIKSCMEWLETQKQASVAFVSFG 283
QY 289 SLLRLPKPQMEETAIIGLSDTKSPVLWVIRRNNEEGDEQAEKEEELKLSFFDRHGTGLRGK 348
Db 284 SFGILFEKOLAEVAIALQESDLNPLWVKEAHIAKLPEGFVESTK-----DR-----AL 332
QY 349 IVTWCSDQLDVLTHKSCVFTVHCWNSAISLACGVPVVCFFQWFDQGTNAKMIEDVWRS 408
Db 333 LVSCNOLVLAHESIGCFTHCGWNSLTALSGFVPMVANAALWTDQPTNARNVELANGAG 392
QY 409 GVRVNRVEGG---VVDREIKRVCVSEVIK---SRELRESAMMKWGLAKEAMDEERGSSMN 463
Db 393 GYRAK--EAGEVIVKSEELVRLKGVMEGESSVKIRESSKKWKDLAYKAN--SEGGSSDR 449
QY 464 NLKNEI 469

Db 450 SINEFI 455
RESULT 10
T00511
Indole-3-acetate beta-glucosyltransferase homolog T20D16.16 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00511
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: Z14159
A:Accession: T00511
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-453 <R0U>
A:Cross-references: EMBL:AC003991; NID:g2642427; PID:g2642442
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 266/2
A:Note: T20D16.16
C:Superfamily: flavonol O3-glucosyltransferase
Query Match 23.9%; Score 593.5; DB 2; Length 453;
Best Local Similarity 33.2%; Pred. No. 5.4e-37;
Matches 159; Conservative 80; Mismatches 181; Indels 59; Gaps 18;
QY 6 HILLATFPAQGHINPSLEFAKRLNTGVDQVTFEVSVALRRMRFETD--PSSRIDFV 62
Db 10 HVMVALPFGHLPNMLKFAKLARTN----LHFTLATIESARDLSLSDPHSLVDLV 64
QY 63 AXXDYDDGLKKGD--DGKNYMSRMKRGTALKDPTLIKLNDAAMGSECYNRVSVVYSH 120
Db 65 ----PFSDCLPKDDPROHEPLETSLRKVGANNFSKII-----EGKRFDCIISVPT----- 111
QY 121 LFSNAAEVAREVDVPSALLWTEPATVDVYFYF---NGVADDIDAGSDQIOLPNLPOLS 177
Db 112 --PWVPAVAANAHPICAILWIEACAGFSVYRYMTNSF--PDLEDPNOKVELPLGFLPLE 168
QY 178 KODLSFLLPSPARFRTLMKEFTLDKPKAKVLINTFDALETEQLKAI--DRYELISI 236
Db 169 VRDLPTLPLSHGAIFNTLMAEFVECL--KDVK--WLANSPYELESVIESMFDLKIPI 226
QY 237 GLPISSIFSDDNDPSSSNKSYGGDLPRKADETYMWLNKSPSSVYVYSGSLRLPKP 296
Db 227 GLVSPFLLGDADEKILDKSL--DMW--KADDYCMEMLDKQV-----SILKSSN 273
QY 297 OMEETAIIGLSDTKSPVLWVIRRNNEEGDEQAEKEEELKLSFFDRHGTGLRGKIVTWCSD 356
Db 274 QVETIATALKNRGVPLFWVIRPKEAENVLDMEV-----EGGVVIEWGOOE 323
QY 357 DVLTHKSYGCVFTHCGWNSAISLACGVPVVCFFQWFDQGTNAKMIEDVWRSVVRVNE 416
Db 324 KILCHMAISCFTVHCWNSLTALSGFVPMVANAALWTDQPTNARNVELANGAG 383
QY 417 EGVVYDRREIKRVCVSEVIK---SRELRESAMMKWGLAKEAMDEERGSSMNLKNEITRI 472
Db 384 VDGLKVAEVERCIDAVTKGTDAADMRRRAELKQATRSAM--APGGLARNLDLFINDI 441
RESULT 11
F71419
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: F71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chawatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113
A:Accession: F71419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <BE>
A:Cross-references: GB:297339; NID:G2244901; PID:e326932; PID:G2244907
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 23.8%; Score 590.5; DB 2; Length 475;
Best Local Similarity 32.5%; Pred. No. 9.7e-37;
Matches 162; Conservative 84; Mismatches 180; Indels 73; Gaps 19;
Qy 6 HILLATPFAQGHINPSLEFAKRLNTGYVDQVTFSTSYAL-RRMRPFTDPSS-----57
Db 9 HVMVSPFGQGHISPLRLGKIASKGLI--VTFVTTEELPKMKRQANNIQDGLKPVG 66
Qy 58 ----RIDFVAXDSDYDGLKGDCKNYSMEKRGTKALKDTLIKLNDAAMSECYNRV 113
Db 67 LGFLREF-----FDGVRVYKEDFDLQKSLVSGKREIKNLVKYKQPV--RCLINN 118
Qy 114 SFVYSHLFSMAEAREVDVPSALLWIEPATVFDVYFYPNGYADIDAGSDQI--QLP 171
Db 119 AFV-----PWVCDIAEELQIPSAVLWVQSCACLAAYVHHVHOLVKKPTEPEITVDVP 172
Qy 172 NLQLSKOD-LPSFLPSSP-ARFTLKKKEDFDLQKSLVSGKREIKNLVKYKQPV--RCLINN 118
Db 173 FKPLTKHDEIFSLPSSPLSSIGTLEQIKRLK--PFSVLITFTFLEKDTIDHMS 230
Qy 226 KATDRYELISGLP--IPSSIFSDGNDPSSSNKSYGGDLFRKADETMDLNSKPSSVY 283
Db 231 QLCPOVNFNPIGPLFTMAKTIRSD-----IKGDI-SKPDSDCIEWLSREFSVV 279
Qy 284 YVSGSLLLPKPMEEIAGLSLDSKPSVLVIRNEEGDQV-----QAEKEKLLSFFD 339
Db 280 YISFGTLAFKQNGIDEIANGLSGLSLVLRPPLEGLAIEPHVLPLELEK-----333
Qy 340 RHGTERLGIKVTWCOLDVTHKSYGCVFTHCGWNSAIESLACGVPVVCPPQWFDQGTNA 399
Db 334 -----GKIVWCQOEKVLAPAVACFLSHCGWNSMTAELTSCVPVVICFPQWGDQVTA 386
Qy 400 KMIEDYVRSQVRV-RVNEEGGVVDRREIKRCVSEVI---KSLRESAMMKGLAKEAMD 455
Db 387 VYMIDVFKTLRSLRGASDERIVPREEVAERLELATVGEKAVELREARNRKEEASAV- 445
Qy 456 EERGSSNNLKNFTRIIN 474
Db 446 AYGTSERNFQEVDRKLV 464

RESULT 12
T00584
Indole-3-acetate beta-glucosyltransferase homolog T27E13.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00584
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.
A:Reference number: Z14177
A:Accession: T00584
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-455 <ROU>
A:Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150406

A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 166/1
A:Note: T27E13.12
C:Superfamily: flavonol O3-glucosyltransferase
Query Match 20.1%; Score 499; DB 2; Length 455;
Best Local Similarity 31.2%; Pred. No. 6.8e-30;
Matches 153; Conservative 74; Mismatches 198; Indels 66; Gaps 17;
Qy 5 RHILLATPFAQGHINPSLEFAKRLNTGYVDQVTFSTSYALRRMRPFTDPSSRIDVAX 64
Db 12 RHVVAMPYPGGRGHINPMNLCRLVRRYVNLHVTFVTEENLGFIGDPKPK-DRHFE STL 70
Qy 65 XDSYDGLKKGDDCKNYSMEKRGTKALKDTLIKLNDAAMSECYNRVSVYSHLF-S 123
Db 71 PNLPSLVRKADFIGFIDAVYTRLEPEPEKLLDSLNSPPPS-----VIPADTYVI 121
Qy 124 WAAEAREVDVPSALLWIEPATVFDVY----YFENGYADIDAGSDQI--QLPNLPQLS 177
Db 122 WAVRVGKRNIPIVSVLWTSATILSFLHSDLLISHGHA-LFEPSEEEVDVYVGLSPTK 180
Qy 178 KQDLPSFLPSSPARFTLKKKEDFDLQKSLVSGKREIKNLVKYKQPV--RCLINN 118
Db 181 LRDLPPFDGYSRVEKT-AKCFDEL---PGARSLFT-TAYELEH-KAIDAFSTKLDI 234
Qy 232 ELISIGLIPSSIFSDGNDPSSSNKSYGGDLFRKADETMDLNSKPSSVYVSGSL 291
Db 235 PVVAIGLPIPFELSVDNDKPN-----YIQWLEEQPEGSVLXISQGSFL 280
Qy 292 RLKPKQMEETAGLSLDSKPSVLVIRNEEGDQV-----QAEKEKLLSFFD RGTGRLGIKVT 351
Db 281 SVSEAOMEELVKLSGVRFLVAKH---GGELKLEALE-----GSLGVVVS 325
Qy 352 WCSOLDVTHKSYGCVFTHCGWNSAIESLACGVPVVCPPQWFDQGTNAKMIEDYVRSQVR 411
Db 326 WCDQLRVLCBKAVGGFTWCGFNSTLEGISYGVPMPLAFPLFMDQLLNKAKMIVEDMRVGM 385
Qy 412 V-RVNEEGGVVDRREIKRCVS-----EVTKSRLRESAMMKGLAKEAMD EERGSSNNL 465
Db 386 IERTKNELLIGREEIKVKKRPMORESEBGEKMRACDLSLSEIRGAV-AKSGSSNNVI 444
Qy 466 KNFTIRIIN 476
Db 445 DEFVRHITN 455
RESULT 13
T45603
glucosyltransferase-like protein - Arabidopsis thaliana
N:Alternate names: protein F12A12.180
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C:Accession: T45603
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23008
A:Accession: T45603
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <CHO>
A:Cross-references: EMBL:AL13314
A:Experimental source: cultivar Columbia; BAC clone F12A12
C:Genetics:
A:Map position: 3
A:Introns: 161/1
A:Note: F12A12.180
C:Superfamily: flavonol O3-glucosyltransferase
Query Match 19.8%; Score 491.5; DB 2; Length 453;

Db 66 --LPSPDFEDLG-----PIEF LHKLNKCEQVSKDCIGQLL-----LQGNIEIACVVD 111

QY 120 HLFSAAEVAREVDVPSALLWIEPATVFDVYFYFNGVADDI-----DAGSDQIQIQLPNL 173

Db 113 EFMYFAEAAAEKFLPNVIFSTTATAFVCRSAFDKLYANSILTPLEKPGQGNELVPEF 172

QY 174 POLSKQDLPSPFLPSSPAREPTLMKKKFTDLQKEPAKVLINTFDALLETQOLKAIDRYEL 233

Db 173 HFLRCKDFP---VSHWASLESMEIYRNTVDKRTASSVIINTASCLESLSRLQOQOL 228

QY 234 IGIGLIPSSIFSDGNDP--SSSKNSYCGDLFRKADETYMDLNSKPSSVYVYFSGSLLR 292

Db 229 IPYIGPLHLVASNSTSLLEENK-----CIEWLNKOKNSVIFVSLGSLAL 276

QY 293 LPKPOMEIATGLSDTKSPVLWIRNE-EGDE--OEOAEZEKKLFFDRHGHETERLGKI 349

Db 277 MEINVEITAGLSDSSKQQLWIRPGSVRGSEWENTLPKFSEKIIS-----GRGYI 328

QY 350 VTWCQOLDVLTHKSGCFVTHCGNSAIBSLACGVPVYCFQWFDQGTNAKMIEDVYRSG 409

Db 329 VKNAPQKEVLHPAVGGFWHCGNNTSLSIGEGVPMICKPFSSDQWVNARYLECVWKIG 388

QY 410 VYRVYNEGGVYDRREIKRVCSEYIKSRELPSAMWKGAKAEAMDEER-----GS 460

Db 389 IOVE-----GDLDRGAVRAVRLWVEEE-----GEGMKRAISLKEQLRASVYSGGS 436

QY 461 SMNNLNKFI 469

Db 437 SHNSLEEFV 445

RESULT 15

T45602

glucosyltransferase-like protein - Arabidopsis thaliana

N;Alternate names: protein F12A12.170

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000

C;Accession: T45602

R;Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artig

submitted to the Protein Sequence Database, December 1999

A;Reference number: Z23008

A;Accession: T45602

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-438 <CHO>

A;Cross-references: EMBL:AL133314

A;Experimental source: cultivar Columbia; BAC clone F12A12

C;Genetics:

A;Map position: 3

A;Introns: 145/1

A;Note: F12A12.170

C;Superfamily: flavonol O3-glucosyltransferase

```

A:Residues: 1-451 SCOP2
A:Cross-references: EMBL:AL133314
A:Experimental source: cultivar Columbia; BAC clone F12A12
C:Genetics:
A:Map position: 3
A:Introns: 160/1
A:Note: F12A12.190
C:Superfamily: flavonol O3-glucosyltransferase

Query Match      18.4%; Score 456.5; DB 2: Length 451;
Best Local Similarity 28.4%; Pred.No. 1e-26;
Matches 139; Conservative 84; Mismatches 191; Indels 75; Gaps 15;

Qy 5 RHLLATEPAQGHINPSLEFAKRLNTQY--VDQVTF--FTSVYALRRMRKETQPSRI 59
   |::| |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 8 RRVLVANPAQGHISPIQLAKNTLHLKGFSTIAQTRFNFTSPSDDETQGFVTIPES-- 65
   |::| |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 60 DFVAXXSDYDDGLKGGDGGKMYSEMRRKRGTKALKLDTLKLNDAAAGSGECYNRVSPVYYS 119
   |::| |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```
QY 228 IDRYELISIGPLIPSSIFSDGNDPSSSNKSYGGDLFRKADETYMDLWLNKSPSSVYVVSF 287
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
207 LEQVGISYVPLGP--LHMTDSSPSS-----LLEEDRSCIEWLNKQPKSVIYISI 255
QY 288 GSLRLPKPQMEIEAIGLSDTKSPVLWVIRRE---EGDEQEAEEEEKLLSPFDRHGTE 344
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
256 GTLGOMETKEVLEMSNGLCNSNQPLWVIRAGSILGTNGIESLPEDVKNKVS-----E 308
QY 345 RLKGIYTWCSQLDVLTHKSVGCFVTHCGWNSAIESLAGVVPVVCFFQWFDQGTNAKMIED 404
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
309 R-GYIVKRAPQIEVLGHPAVGGFWSHCCGWSILESIGEGVPMICKPFFHGEQKLNAMYIES 367
QY 405 VWRSGVVRVNEEGGVDRRETKRCVSEVKSRE---LRESAMMWKGLAKEAMDEERGSS 461
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
368 VWRIGFQVE-----GKYDRGEVERAVKRLIYDDEGAGWRERALVLKEKIKASV-SSGGAS 421
QY 462 MNMLKNFI 469
Db   : : : :
422 YNALEDIV 429
```

Search completed: August 1, 2000, 21:27:47
Job time: 11215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:35 ; Search time 3616.02 Seconds
(Without alignments)
1752.438 Million cell updates/sec

Title: US-09-147-955-9
Perfect score: 1437
Sequence: 1 ttcaaacatacaactgatat.....tttatttgagtaaaaaaaa 1437

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues 10495684
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	EST:*
1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*
44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_est45:*
80:	gb_est46:*
81:	gb_est47:*
82:	em_est35:*
83:	em_est36:*
84:	em_est37:*
85:	gb_est48:*
86:	gb_est49:*
87:	gb_est50:*
88:	gb_est51:*
89:	gb_est52:*
90:	gb_est53:*
91:	gb_est54:*
92:	gb_est55:*
93:	gb_gss1:*
94:	gb_gss2:*
95:	gb_gss3:*
96:	gb_gss4:*
97:	em_gss1:*
98:	em_gss2:*
99:	em_gss3:*
100:	em_gss4:*
101:	gb_gss5:*
102:	gb_gss6:*
103:	gb_gss7:*
104:	gb_gss8:*
105:	gb_gss9:*
106:	em_gss5:*
107:	em_gss6:*
108:	em_gss7:*
109:	em_gss8:*
110:	em_gss9:*
111:	em_gss10:*
112:	em_gss11:*
113:	gb_gss10:*
114:	gb_gss11:*
115:	em_gss12:*
116:	gb_gss12:*

```

117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	210.8	14.7	720	79	AW622172
2	101.4	7.1	459	81	C72454
3	100.4	7.0	604	46	A1973593
4	98.4	6.8	428	43	A1737582
5	94	6.5	492	47	A1996331
6	93.2	6.5	444	88	R64982
7	70.2	4.9	409	31	AA842442
8	64.4	4.5	960	48	AU066704
9	63.4	4.4	367	81	D24856
10	58.6	4.1	643	74	AW600034
11	55.4	3.9	1001	91	W41743
12	53	3.7	488	91	W70677
13	51.6	3.6	522	79	AW657202
14	51.2	3.6	388	71	AW346140
15	50	3.5	488	72	AW410589
16	49.4	3.4	1101	122	CNS0039G
17	48.4	3.4	1101	122	CNS010CR
18	46	3.2	1101	123	CNS016HF
19	44.4	3.1	928	122	CNS00DKY
20	43.8	3.0	570	64	AW153633
21	43.8	3.0	1101	122	CNS008U5
22	43.2	3.0	951	122	CNS004TM
23	43	3.0	490	93	AQ028894
24	42.6	3.0	997	123	CNS005TE
25	42.4	3.0	979	123	CNS0161W
26	42.2	2.9	566	43	A1722252
27	42	2.9	288	71	AW359817
28	41.6	2.9	1100	123	CNS016Q4
29	41.6	2.9	1101	122	CNS00E5I
30	41.4	2.9	1101	123	CNS017MC
31	41.4	2.9	1101	123	CNS017PL
32	41.2	2.9	952	123	CNS0172P
33	41.2	2.9	1101	122	CNS00E7W
34	40.6	2.8	589	87	N96439
35	40.6	2.8	822	122	CNS009CW
36	40.6	2.8	996	123	CNS01846
37	40.6	2.8	1204	123	CNS016E2
38	40.4	2.8	827	122	CNS00EU3
39	40.2	2.8	1101	122	CNS0039A
40	40	2.8	997	122	CNS005TE
41	40	2.8	1027	123	CNS015PL
42	40	2.8	1101	122	CNS0039L
43	39.6	2.8	979	123	CNS0161W
44	39.6	2.8	1043	123	CNS0145P
45	39.6	2.8	1101	122	CNS000D1
					AW622172 EST3112970
					C72454 C72454 Rice
					A1973593 sc88h08.y
					A1737582 T110050.e
					A1996331 T01664961
					R64982 13486 Lambd
					AA842442 MBAFCW8D0
					AU066704 AU066704
					D24856 SMC2662A.F
					AW600034 R09vAFACP
					W41743 mc63a08.r1
					W70677 me42c04.r1
					AW657202 109900.MA
					AW346140 27337.MAR
					AW410589 fh07c02.x
					AL063921 Drosophil
					AL098805 Drosophil
					AL0106749 Drosophil
					AL017865 Drosophil
					AW153633 fd23h11.y
					AL026640 Drosophil
					AL055719 Drosophil
					AQ028894 CIT-HSP-2
					AL060767 Drosophil
					AL106190 Drosophil
					A1722252 fd23h12.y
					AW359817 46864.MAR
					AL107062 Drosophil
					AL069797 Drosophil
					AL108222 Drosophil
					AL108339 Drosophil
					AL108703 Drosophil
					AL069847 Drosophil
					N96439 21018.CD4-1
					AL053618 Drosophil
					AL108864 Drosophil
					AL106628 Drosophil
					AL069854 Drosophil
					AL063915 Drosophil

ALIGNMENTS

RESULT	1
AWG22172	720 bp mRNA EST 28-MAR-2000
LOCUS	EST1312970 tomato root during/after fruit set, Cornell University
DEFINITION	Lycopersicon esculentum cDNA clone CLEX14O10.5', mRNA sequence.

Db 188 CCTGTGTCGATTATCTTACATGGATATTTTTGAAGACCAAGTATGGCGTCAGGAAGTTC 247
Qy 702 gtccgtcctcatattgtagggagcctcctcctcgcttccttccgactcaggagtc 761
Db 248 ATCGGTGTGGAGTTTGCGTGCCTGGACTATATATTAGTAGTATTTCAGATGTCCATGCC 307
Qy 762 actgcot---gctggttcgaatcctcttttggtgattcttcttgtcatacaggctctat 817
Db 308 TCTGATCGAGCTAGACCACCCCTTTGAAGGGAGATTTCGCTGTAATCTTTGGGTCCAT 366

RESULT 3
LOCUS A1973593 604 bp mRNA EST 18-APR-2000
DEFINITION SC89H08.Y1 Gm-cl019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl019-472 5' similar to TR:059785 Q59785 HYPOTHETICAL 56.8 KD
PROTEIN. ; mRNA sequence.
ACCESSION A1973593
VERSION A1973593
KEYWORDS GI:5770419
SOURCE EST.
ORGANISM Glycine max
Eukarya; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 604)
Shoenmaker,R., Keim,P., Vodkin,L., Erpelnding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson.R.
Public Soybean EST Project
Unpublished (1999)

TITLE On Aug 21, 1998 this sequence version replaced gi:3704615.
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1451 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 436.
Location/Qualifiers
1..604
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl019-472"
/clone_lib="Gm-cl019"
/tissue_type="immature seed coats of greenhouse grown plants"
/lab_host="DH10B (Gibco BRL)"
/note="vector: pSPORT1 (Life Technologies); Site_1: Not I; site 2: Sal I; this cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies Superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was

Contact: Blaxter ML
Institute of Cell Animal and Population Biology
University of Edinburgh
Aspathor Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Tel: +44 131 670 5450
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at
<http://www.sanger.ac.uk/bugia/AF/MBAFWC8D08T3.html>
Seq primer: T3

5', mRNA sequence.
AU066704
AU066704.1 GI:4967441
EST.
house mouse.
Mus musculus

REFERENCE

O'Grady, L.: 1987, 'The Evolutionary History of Vertebrates', Oxford University Press, New York.

AUTHORS

Kanayama, Y., Hirata, M., Tanuma, R., Ito, A., Hashimoto, K., Kusuda, J., Toyoda, A., Suzuki, Y., Sasaki, M., and Sugano, S.

TITLE

Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method

JOURNAL

Unpublished (1999)

JOURNAL
COMMENT
On Apr 7, 1998 this sequence version replaced q13036257.
Unpublished (1999)
ribulose made by C4/C3 crassulacean mesophyll

Contact: Katsuyuki Hashimoto
Division of Genetic Resources
National Institute of Infectious Diseases
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo
Email: khashi@nih.go.jp
URL: <http://www.nih.go.jp/voken/genbank/>.

FEATURES	SOURCE
-----------------	---------------

Xho I; Lymphatic filarial nematode parasite of humans. mRNA was prepared from approximately 50 adult females isolated from the peritoneal cavity of jirds and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 5 x 10⁶ independent recombinants and the average insert size is ~900bp. The library was constructed by Michelle Lottotte-Waniewski. The library is available from Dr. S.A. Williams, email: williams@uakron.edu.

```

location=Vdairr185
i..960
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCB-1380"
/clone_lib="Sugano mouse brain mncb"
/sex="female"
/dev_stage="adult"
/lab_host="TOP10"
/note="Organ: brain; Vector: pME185-
was primed with an oligo(dT) primer
[ATGCGCCCTTTTCTTTTTTTTTTTTT] double-a
and cloned into distinct DraIII sites
XhoI sites just outside the DraIII s
isolate the cDNA insert. Size select
exclude fragments <1.5 kb. Library w
Sugano et al. (University of Tokyo,
Science). Custom primer for sequenc
[CTTCGTGCTAAAGCTGCCT].

```

BASE COUNT	104 a	60 c	86 g	159 t
ORIGIN				

Query Match 4.9%; Score 70.2; DB 31; Length 409;
Best Local Similarity 50.4%; Pred. NO. 1.2e-08;
Matches 205; Conservative 0; Mismatches 193; Indels 9

[illegible]

BASE COUNT	250 a	215 c	214 g	230 t	51 others
ORIGIN					
Query Match		4.5%	Score 64.4	DB 48	Length 960
Best Local Similarity		48.6%	pred. No. 5.7e-07		
Matches 176:	Conservative	0	Mismatches 186	Indels	Gaps 0

Qy	778	atcctcttcttggtgattttctctgtcataaacaggctctctatttggttccactcagcagcgtg	837
Db	30	ATGTTCTTATTTGGTGCATCTTATGTCCTGCTTGGGGCTTCCCTGTATGCTGTGTCATAACG	89
Qy	838	tuggtcaggaataactgctggaagaggaagatcgtattgaagtagcagcaatgatacgggtg	897
Db	90	TGTGTGAAGATACATCGTGAGAGAGCTGACGACAGGAAATTTTAGGAATGTGTGGAC	149
Qy	898	tatttggtagtctcaatcaagtgcgaacaggagatactgtgctggagaggaatgcctctctcat	957
Db	150	TATTTGGAACAATCATACGGCATACAGCTATTGTATGTGGAATATAAAGACATATGGCC	209
Qy	958	caatgcagtggtctactggaacttttggcagcctatgttttattgcaactgtccagcttcc	1017
Db	210	GGATTCAGTGGGACTGGMAAATTGCCCTGTTATTGTGGCAATTTGCCCTCTATATGTTTT	269
Qy	1018	tctctgcacactcaaccccttttctctcaagatgagtgcgctgcatcttttccaactcttt	1077
Db	270	GCCTGTACAGCTTCAATCGCTTGGTGATTTAAAGTCACAGATGCCACTTCAAGTCAACCTGG	329
Qy	1078	ccatgcttcaactctgatatgtggctgttgcaattaggacattcatatacaaccaggagg	1137
Db	330	GCATCCTGCACAGCTGACCTTTACAGCCTTTCTTTGGGCTTTTTTTATTCGAGTATAAAG	389

RESULT	8
LOCUS	AU066704
DEFINITION	Sugano mouse brain mcb Mus musculus EST
DEFINITION	AU066704 Sugano mouse brain mcb Mus musculus cDNA clone MNCB-1380
DEFINITION	AU066704 Sugano mouse brain mcb Mus musculus cDNA clone MNCB-1380

SNOVAFCAP42G12SK Onchocerca volvulus adult female cDNA
 (SAW98MLM-OVAF) Onchocerca volvulus cDNA clone SNOVAFCAP42G12 5',
 mRNA sequence.
 AW600034
 AW600034.1 GI:7287547
 EST.
 Onchocerca volvulus.
 Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 1 (bases 1 to 643)
 Lizotte-Waniewski, M., and Williams, S.A.
 Genes expressed in adult female stage of Onchocerca volvulus
 Unpublished (1998)
 On Aug 21, 1998 this sequence version replaced gi:3705002.
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.
 Location/Qualifiers
 1..643
 /organism="Onchocerca volvulus"
 /db_xref="taxon:6282"
 /clone="SNOVAFCAP42G12"
 /clone.lib="Onchocerca volvulus adult female cDNA
 (SAW98MLM-OVAF)"
 /sex="female"
 /dev_stage="adult"
 /lab_host="Xfil-Blue MRF"
 /note="vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. Two adult
 female worms of Onchocerca volvulus were isolated from
 consenting patients and quick frozen. Adult female mRNA
 was converted to double-stranded cDNA using reverse
 transcriptase and oligo(dT) followed by RNase H and DNA
 pol I. The library has 7 x 10E5 independent recombinants
 and the average insert size is ~1100bp. The library was
 constructed by Michelle Lizotte-Waniewski with worms
 provided by Dr. Sara Lustigman. The library is available
 from Dr. Steven A. Williams, email: genome@smith.edu."
 200 a 112 c 123 g 208 t
 BASE COUNT
 ORIGIN
 Query Match 4.1%; Score 58.6; DB 74; Length 643;
 Best Local Similarity 50.9%; Pred. No. 2e-05;
 Matches 139; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
 Qy 938 ggagaggaatgccctccatcaatgcagtgctactggacttttggcgcctatgtt 997
 Db 1 GGAGCATCATCATAGTTGGCGATCGGTCAATGTGTGAGAACATCATAGTATTCATCCTCAT 60
 Qy 998 ttatgcactgtccagcttctctctgtcacactcacccctttctctcctcaagtgaagtgg 1057
 Db 61 ATTTATTGGCTCTATGTTCTTCGTTCTATTTCATCGGTGAGCGTAGTCTTAGAGAGTCTTC 120
 Qy 1058 cgtgcatttttccaatctttccatcgtctaccatctgatatgctggtgctgttgcattaggac 1117
 Db 121 CGCTTCTATGTTTCAACTATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 180
 Qy 1118 attcatatacaaacaggaggttgattggtttactattttggccttttctcgttgttgc 1177
 Db 181 ATTCATGTTCAATTATCAATTTTCATGAATGTTGATTTGTCATGCTTTTGACCTAGTAAATAT 240
 Qy 1178 tggatattcatcatatacaaaaaacagagaagga 1210
 Db 241 CGGTTCTCTTAATATATTTCCATCCGGAACAACAGA 273

```

RESULT 11
W41743      1001 bp      mRNA      EST      20-MAY-1996
DEFINITION  mc63a08.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:353174 5', mRNA sequence.
ACCESSION  W41743
VERSION    W41743.1 GI:1325772
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 1001)
AUTHORS   Maria.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,
Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M.,
Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B.,
Theising.B., Wylie.T., Lennon.G., Soares.B., Wilson.R. and
Waterston.R.
TITLE     The WashU-HMMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:635814.
Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:224974
Seq primer: ETPrimer
High quality sequence stop: 409.
Location/Qualifiers
1..1001
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:353174"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DHI08"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTCAAGTCGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'] on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 202 a 200 c 226 g 265 t 108 others
ORIGIN
Query Match 3.9%; Score 55.4; DB 91; Length 1001;
Best Local Similarity 47.4%; Pred. No. 0.00017;
Matches 207; Conservative 0; Mismatches 221; Indels 9; Gaps 1;
QY 600 acatgattacagtgagcactaggttggtgacacatggtggtccatctctt 659
DB 1 AGCAGCTCTGACCATGTGCCAGCTCTTGATGTGTTGGGATTCCTGTGTGATGGCTCTC 60
QY 660 acatgagttctcaggacacaaatactctgtataccagtttgctggtcgtctattgt 719
DB 61 TCGTGGTTATTCCTCGGGAAGATCAAAAGTATCCACTTCATTCCTGCTGTGCTGTGT 120

```

```

QY 720 gtaggaggcctcctcctcctcgtgctcttcttcctcgaactca-----ggggctcactcgtcgt 770
DB 121 CTGCTGGGTGTAGGACCATGGTTGGTGCAGACATATTAGCTGGCGGAGAGATAATTTCA 180
QY 771 ggttcgaatcctctcttgggggatttcttctgtcataacaggctctattttgttcacactc 830
DB 181 GGGAGTGTATGTTCTTATTGTGTGACATCTTAGTCTCTGCTTGGGGCTTCCCTGTATGCTGTG 240
QY 831 agcactgttggtcaggaatactcgtgtaagagaaagatacgtattgaagtagtaagaatg 890
DB 241 TCTAACGTGTGTGAAGAAATACATCGTGAAGAAGCTGAGCAGACAGGAATTTTAGGAATG 300
QY 891 atcgggtatttgggtatgctcactcagtcgcaaccgagatctactgtgctgtagagggaatgcc 950
DB 301 GTGGGACATATTGGGAACAATCATCAGCGGCATACACGCTATTGATTGTGGGAATATAAGAC 360
QY 951 ctctcatcaatgcagtaggttctactggacttttggagcctatgttcttatcagcactgccc 1010
DB 361 ATGCCCGGATTCAGTGGGACTGGAAAAATTCGCCCTGTTATTGTGGCATTTGCCCTCTGT 420
QY 1011 agcttcctctctctgcac 1027
DB 421 ATGTTTTCCTGTACAC 437
RESULT 12
W70677      488 bp      mRNA      EST      17-JUN-1996
DEFINITION  mc42c04.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:390150 5', mRNA sequence.
ACCESSION  W70677
VERSION    W70677.1 GI:1380829
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 488)
AUTHORS   Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,
Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M.,
Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B.,
Theising.B., Wylie.T., Lennon.G., Soares.B., Wilson.R. and
Waterston.R.
TITLE     The WashU-HMMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1290773.
Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:241982
Seq primer: ETPrimer
High quality sequence stop: 340.
Location/Qualifiers
1..488
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:390150"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DHI08"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTCAAGTCGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'] on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 202 a 200 c 226 g 265 t 108 others
ORIGIN
Query Match 3.9%; Score 55.4; DB 91; Length 1001;
Best Local Similarity 47.4%; Pred. No. 0.00017;
Matches 207; Conservative 0; Mismatches 221; Indels 9; Gaps 1;
QY 600 acatgattacagtgagcactaggttggtgacacatggtggtccatctctt 659
DB 1 AGCAGCTCTGACCATGTGCCAGCTCTTGATGTGTTGGGATTCCTGTGTGATGGCTCTC 60
QY 660 acatgagttctcaggacacaaatactctgtataccagtttgctggtcgtctattgt 719
DB 61 TCGTGGTTATTCCTCGGGAAGATCAAAAGTATCCACTTCATTCCTGCTGTGCTGTGT 120

```


TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle

JOURNAL Unpublished (2000)

COMMENT On Jan 6, 2000 this sequence version replaced gi:66751137.

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 20 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGAGC
Plate: 18 row: 0 column: 24
Seq primer: ATTTAGGTACACATAG.

FEATURES
Location/Qualifiers
1..383
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4B0V"
/tissue_type="pooled"
/lab_host="PH103"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 85 a 90 c 93 g 115 t

ORIGIN

Query Match 3.6%; Score 51.2; DB 71; Length 383;
Best Local Similarity 47.0%; Pred. No. 0.0018;
Matches 158; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 857 gaagagaaagatcgtattgaagtagcaatgacatgctggtgattgttgatgctatcag 916
Db 2 GAAGAAGCTGACGAGAAAGGAGTATTTAGACTGGTGGCTTATTTGGAAACAATTATCAG 61

Qy 917 tgaacaggagattactgtctgagaggaatgccctctcatcaatgcagtggtctactgg 976
Db 62 TGGCATACAGCTATTGATTGGAATATAGGATATTGCCAGCATTCCTGGACTGGAA 121

Qy 977 acttttggagccatagtgttttattgacatgctccagcttctctctgcacatccccc 1036
Db 122 AATTGCCCTGCTATTTGTAGCGTGTTCCTCTGTATGTTTGGCTGTATAGCTTTCATGCC 181

Qy 1037 ttcttctcaagatgagtgagcgctgcatcttttcaatcttccatgcttcatctgatat 1096
Db 182 ACTGGTATCAAGTCACCAAGTCCACTTCTGTCAACCTGGGCATCTTCAGAGTGACCT 241

Qy 1097 gtgggctgttgcattaggacattcatatacaaccaggaggtgtgattggttatactatt 1156
Db 242 CTACAGCTCTTTCTTCGGAGCTCTTCTGTGTTGGCTATAAGTGTTCAGGACTGTACATCCT 301

Qy 1157 ggccttttctcgtgtgtgttgggaattcatata 1192
Db 302 GTCCTTCACCGTCATCATGTTGGGGTTCATCTCTGA 337

RESULT 15

AW410589 488 bp mRNA EST 16-FEB-2000

LOCUS fh07c02.x1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2961410 5', mRNA sequence.

DEFINITION

ACCESSION AW410589.1 GI:6936130

VERSION

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1. (bases 1 to 488)
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Jan 19, 1998 this sequence version replaced gi:2284719.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Plate: LHAM56 row: F column: 3
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers
1..488
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2961410"
/clone_lib="NIH_MGC_17"
/tissue_type="thadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES
source

BASE COUNT 120 a 99 c 126 g 143 t

ORIGIN

Query Match 3.5%; Score 50; DB 72; Length 488;
Best Local Similarity 51.3%; Pred. No. 0.0041;
Matches 116; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 527 atggtactggtatgctctctctggggttcgtggagctccacggcaattatcttggtaataa 586
Db 178 ATGGTGGGAAGTACATCTCTGCTGGGACTAGCAGATGTGGAAGCTAATTTATGTGATCGTCAG 237

Qy 587 agcatcgagttgacatcgattacgagtgtagcactactggtgattgttggacaactcgtgtg 646
Db 238 AGCCTACCAAGTACACAACCTCTAACCAAGTGTCCAGCTTTTGGATTGCTTTGGGATTCCCTGT 297

Qy 647 gtccatcatctttcacatggatgttctctaggcaacaaatactctgtataccaggttgctgg 706
Db 298 GTTGATGGCTCTGTGCTGGTTTATCTTCATGCAAGATACAGAGTGATCCACATTCATCGC 357

Qy 707 tgcgtctattgttagaggagccctctctctctgctgtctcttcttcgcac 752
Db 358 CGTGGCTCTGCTGTGTTGGGTGTAGGAACCACTGGTTGTGTGCAGAC 403

Search completed: August 1, 2000, 14:38:44
Job time: 62716 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:18:36 ; Search time 152.01 seconds
(without alignments)
1300.360 Million cell updates/sec

Title: US-09-147-955-9
Perfect score: 1437
Sequence: 1 ttcaaaactcataacgtgat.....tttatttgagtaaaaaaa 1437

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2.6/ptodata/2/ina/5C_COMB.seq: *
4: /cgn2.6/ptodata/2/ina/5D_COMB.seq: *
5: /cgn2.6/ptodata/2/ina/6_COMB.seq: *
6: /cgn2.6/ptodata/2/ina/PTCTUS_COMB.seq: *
7: /cgn2.6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	3.1	7218	1	US-08-232-463-14
C 2	43.6	3.0	7218	1	US-08-232-463-14
C 3	35.8	2.5	480	1	US-08-282-561-4
C 4	35.8	2.5	480	1	US-08-550-544-4
5	34.8	2.4	4951	6	PCT-US95-06119-5
6	34.8	2.4	4951	6	PCT-US95-06119-5
7	33	2.3	1297	1	US-08-370-975B-12
8	33	2.3	1297	1	US-08-370-975B-1
9	32.8	2.3	19124	4	US-08-487-826B-13
C 10	32.4	2.3	11225	7	5182210-9
C 11	32.2	2.2	2472	5	US-08-335-844A-7
C 12	32.2	2.2	3358	5	US-08-335-844A-20
C 13	32	2.2	729	7	5182210-21
C 14	32	2.2	5138	3	US-08-476-062A-39
C 15	31.2	2.2	2080	1	US-08-077-939-18
C 16	31.2	2.2	2080	1	US-08-461-599-18
C 17	31.2	2.2	2080	1	US-08-461-621-18
C 18	31.2	2.2	2080	2	US-08-465-334-18
C 19	31	2.2	1298	2	US-08-473-157A-2
C 20	31	2.2	10079	4	US-08-476-866-20
C 21	30.8	2.1	179	5	US-09-065-474-100
C 22	30.8	2.1	179	5	US-09-065-474-100
C 23	30.8	2.1	810	2	US-09-031-485-9
C 24	30.8	2.1	810	2	US-09-031-485-10
C 25	30.8	2.1	810	2	US-08-847-429A-9
C 26	30.8	2.1	810	2	US-08-847-429A-10

27	30.8	2.1	810	5	US-09-065-474-9	Sequence 9, Appli
C 28	30.8	2.1	810	5	US-09-065-474-10	Sequence 10, Appl
29	30.8	2.1	864	5	US-09-065-474-144	Sequence 144, App
C 30	30.8	2.1	864	5	US-09-065-474-146	Sequence 146, App
31	30.8	2.1	936	2	US-09-031-485-4	Sequence 4, Appli
C 32	30.8	2.1	936	2	US-09-031-485-5	Sequence 5, Appli
33	30.8	2.1	936	2	US-08-847-429A-4	Sequence 4, Appli
C 34	30.8	2.1	936	2	US-08-847-429A-5	Sequence 5, Appli
35	30.8	2.1	936	5	US-09-065-474-4	Sequence 4, Appli
C 36	30.8	2.1	936	5	US-09-065-474-5	Sequence 5, Appli
37	30.8	2.1	937	2	US-09-031-485-1	Sequence 1, Appli
C 38	30.8	2.1	937	2	US-09-031-485-3	Sequence 3, Appli
39	30.8	2.1	937	2	US-08-847-429A-1	Sequence 1, Appli
C 40	30.8	2.1	937	2	US-08-847-429A-3	Sequence 3, Appli
41	30.8	2.1	937	5	US-09-065-474-1	Sequence 1, Appli
C 42	30.8	2.1	937	5	US-09-065-474-3	Sequence 3, Appli
43	30.8	2.1	1029	2	US-09-031-485-6	Sequence 6, Appli
C 44	30.8	2.1	1029	2	US-09-031-485-8	Sequence 8, Appli
45	30.8	2.1	1029	2	US-08-847-429A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14


```
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 162
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 356
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
US-08-282-581-4
```

```
Query Match 2.5%; Score 35.8; DB 1; Length 480;
Best Local Similarity 46.1%; Pred. No. 0.14;
Matches 118; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
```

```
QY 1182 atattcatatatacaaaacagagagatcctaacaatagagagcccttgagaatgga 1241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 444 ATATATATATATAAAAAAGACAAAGGAGGTATCTCAGCAATTTCTCACCACAAATTGA 385

QY 1242 aacttgatcatgaatagtcctcttgaggatcaagatgacacacccaagaaccatag 1301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 384 TTACTTTATTAATAAATACTACTCCTTNGCTGTTTGGGACACCAACAAGCTAAAA 325

QY 1302 ctagcttggcccaatctttcatcaacagtttcaaatcttgcgtgaggggagagaga 1361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 324 TCCAACTGACTGCTCTCTTTTATATATATATATATATATATATATATATATATA 265

QY 1362 tcgagatactaataatgacgtctattatagctgagaggtttttttttatttatta 1421
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 264 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

QY 1422 tttagagtaaaaaaaa 1437
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 204 AAATTTTAAAAACACA 189
```

RESULT 4

```
US-08-550-544-4/c
; Sequence 4, Application US/08550544
; Patent No. 5689056
; GENERAL INFORMATION:
; APPLICANT: Cramer, Carole L.
; APPLICANT: Wellenborn, Deborah L.
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/550.544
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/100,816
; FILING DATE: 02-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
```

```
; REGISTRATION NUMBER: 18872
; REFERENCE/DOCKET NUMBER: 7956-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 162
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 356
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x-Unknown nucleotide"
US-08-550-544-4

Query Match 2.5%; Score 35.8; DB 1; Length 480;
Best Local Similarity 46.1%; Pred. No. 0.14;
Matches 118; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
```

```
QY 1182 atattcatatatacaaaacagagagatcctaacaatagagagcccttgagaatgga 1241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 444 ATATATATATATAAAAAAGACAAAGGAGGTATCTCAGCAATTTCTCACCACAAATTGA 385

QY 1242 aacttgatcatgaatagtcctcttgaggatcaagatgacacacccaagaaccatag 1301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 384 TTACTTTATTAATAAATACTACTCCTTNGCTGTTTGGGACACCAACAAGCTAAAA 325

QY 1302 ctagcttggcccaatctttcatcaacagtttcaaatcttgcgtgaggggagagaga 1361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 324 TCCAACTGACTGCTCTCTTTTATATATATATATATATATATATATATATATATA 265

QY 1362 tcgagatactaataatgagcgtctattatagctgagaggtttttttttatttatta 1421
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 264 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

QY 1422 tttagagtaaaaaaaa 1437
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 204 AAATTTTAAAAACACA 189
```

RESULT 5

```
US-08-867-030B-5
; Sequence 5, Application US/08867030B
; Patent No. 5948900
; GENERAL INFORMATION:
; APPLICANT: Yother et al.
; TITLE OF INVENTION: Streptococcus pneumoniae
; TITLE OF INVENTION: Capsular Polysaccharide Genes and Flanking Regions
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benjamin Aaron Adler, Ph.D., J.D.
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch floppy
; OPERATING SYSTEM: Apple Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/867,030B
FILING DATE: June 2, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/243,546
FILING DATE: May 16, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5923
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 777-2321
TELEFAX: (713) 777-6908
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4951 bp
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE:
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
POSITION IN GENOME:
US-08-867-030B-5

Query Match 2.4% Score 34.8; DB 4; Length 4951;

Best Local Similarity 46.3%; Pred. No. 1.2; 132; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1098 tgggtgttgcaattagacattcatatacaaccaggaggttgattggttatactatttg 1157
Db 3540 TTGACAGATGCTATTGTATACATTAAGACACAGAGTCTTTTTCGGCGTGAATTTGTC 3599
QY 1158 gcttttctcgttctgttggtggaattcatatatacaaaacagagagagatcctaac 1217
Db 3600 GCGAACGTTACGATGTGGTGAAGTTAAATTTATGAAACATCAATTGATTATGCT 3659
QY 1218 aatacagagcccttggaacttggaacttggaacttggaacttggaacttggaactcaa 1277
Db 3660 CTTCAACATCTCTCAGATTAAGAGAGGTTAAAAAATTACGTTATTCACCTTGGTAAGCAA 3719
QY 1278 gatgacacacacagaaac 1337
Db 3720 TTGGAGAGCTAGATGACTGTTCGTCAGTGGACACCTATGAAATGTTATAGAAGTTATC 3779
QY 1338 aataat 1343
Db 3780 AAAAAAT 3785

RESULT 6

PCT-US95-06119-5
Sequence 5, Application PC/TUS9506119
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE CAPSULAR
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: PatentIn Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06119
FILING DATE: CONCURRENTLY HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/243,546
FILING DATE: 16-MAY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: AMCY018P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4951 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
PCT-US95-06119-5

Query Match 2.4% Score 34.8; DB 6; Length 4951;

Best Local Similarity 46.3%; Pred. No. 1.2; 132; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1098 tgggtgttgcaattagacattcatatacaaccaggaggttgattggttatactatttg 1157
Db 3540 TTGACAGATGCTATTGTATACATTAAGACACAGAGTCTTTTTCGGCGTGAATTTGTC 3599
QY 1158 gcttttctcgttctgttggtggaattcatatatacaaaacagagagagatcctaac 1217
Db 3600 GCGAACGTTACGATGTGGTGAAGTTAAATTTATGAAACATCAATTGATTATGCT 3659
QY 1218 aatacagagcccttggaacttggaacttggaacttggaacttggaacttggaactcaa 1277
Db 3660 CTTCAACATCTCTCAGATTAAGAGAGGTTAAAAAATTACGTTATTCACCTTGGTAAGCAA 3719
QY 1278 gatgacacacacagaaac 1337
Db 3720 TTGGAGAGCTAGATGACTGTTCGTCAGTGGACACCTATGAAATGTTATAGAAGTTATC 3779
QY 1338 aataat 1343
Db 3780 AAAAAAT 3785

RESULT 7

US-08-370-975B-12
Sequence 12, Application US/08370975B
Patent No. 5622851
GENERAL INFORMATION:
APPLICANT: Mailey, Frank
APPLICANT: Mailey, Gladys F.
APPLICANT: Weiner, Karen X.B.
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timlan, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
; US-08-370-975B-12

Query Match      2.3%; Score 33; DB 1; Length 1297;
Best Local Similarity 51.7%; Pred. No. 1.9;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 71 agcatcatcatgatcaatctctttataatgaggaatgagtaacaaggagtggtttt 130
   || |||| |||| |||| || || || || || || || || || || || || || ||
Db 3 AGTTACATCTCAATTCATCTCCAGAAGATTGGGATTCGTCCTTCTTAAGAGGTGCTAAT 62

QY 131 gttactcagcttcaacctacgtactactactactactactactactactactactact 190
   || || || || || || || || || || || || || || || || || || || ||
Db 63 GCCTTTCATCTTGAAGTTACACATAAATCTTACTAGCCAGTAGTGCAAAAGTAGGCATC 122

QY 191 taataataatggcgatagatctt 215
   || || || || || || || || || || || || || || || || || || || ||
Db 123 TAAAGNATATAAAGCCTCAAAATCTT 147

RESULT 8
US-08-370-975B-1
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timlan, Susan J.
; REGISTRATION NUMBER: 34,103

```

```

; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26764 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
; US-08-370-975B-1

Query Match      2.3%; Score 33; DB 1; Length 26764;
Best Local Similarity 51.7%; Pred. No. 12;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 71 agcatcatcatgatcaatctctttataatgaggaatgagtaacaaggagtggtttt 130
   || |||| |||| |||| || || || || || || || || || || || || || ||
Db 25470 AGTTACATCTCAATTCATCTCCAGAAGATTGGGATTCGTCCTTCTTAAGAGGTGCTAAT 25529

QY 131 gttactcagcttcaacctacgtactactactactactactactactactactactact 190
   || || || || || || || || || || || || || || || || || || || ||
Db 25530 GCCTTTCATCTTGAAGTTACACATAAATCTTACTAGCCAGTAGTGCAAAAGTAGGCATC 25589

QY 191 taataataatggcgatagatctt 215
   || || || || || || || || || || || || || || || || || || || ||
Db 25590 TAAAGNATATAAAGCCTCAAAATCTT 25614

RESULT 9
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs

```



```
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5138 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 95...3604
US-08-476-062A-39

Query Match      2.2%; Score 32; DB 3; Length 5138;
Best Local Similarity 61.7%; Pred. No. 8.9;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 233 agcctgcagggtggttaatttcctcggtgtggaataataataataataatagc 292
    ||||| || | | | ||||| |||| | | | ||||
Db 3475 AGCCTCCATCTCTCTCCCTCAGGTTTGAAGAAACCAACCTTGTACAGCACTATGNA 3416

QY 293 gatgagcagcagcagcagcag 313
    ||||| ||||| ||||| |||||
Db 3415 AATGAGCAGCAGCAGCAGCAG 3395

RESULT 15
US-08-077-939-18/C
; Sequence 18, Application US/08077939
; Patent No. 5521088
; GENERAL INFORMATION:
; APPLICANT: FUJII, Toshio
; APPLICANT: IWAMATSU, Akihiro
; APPLICANT: YOSHIMOTO, Hiroyuki
; APPLICANT: MINETOKI, Toshitaka
; APPLICANT: BOGAKI, Takayuki
; APPLICANT: NAGASAWA, Naoshi
; TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/077,939
; FILING DATE: 18-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 184328/1992
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 629997/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 49441/101 KYPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2080 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 311..1888
US-08-077-939-18

Query Match      2.2%; Score 31.2; DB 1; Length 2080;
Best Local Similarity 48.8%; Pred. NO. 9.1;
Matches 84; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1066 ttttcaatctttccatcttaccatctgatgtggcgctgttgcaattaggacattcatat 1125
    ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 TTCTCTCTGTTTCCATGCTCTCAAGTTTGTGGTGGCTTTAAAAAAGACGCTTCTCGA 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1126 aaaccagcagggttgattggttatctactatttggccttttctctctgttctgttggaatat 1185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 AGAGCCATGAAAAACAATAGAAAAACATATTTTGGACTTCGCTCCAGCGATGCTGATGCCA 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1186 tcatatatacaaaaaaacagagagagatctctaacaatacagagagcccttgagaa 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 GCCCAAAACAAATAGGCGATCAAAAAAATTACAAAAATCACAGCCCTTTAGAA 95

Search completed: August 1, 2000, 18:19:33
Job time: 74422 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:18:09 ; Search time 236.66 Seconds

(without alignments)

1519.167 Million cell updates/sec

Title: US-09-147-955-9

Perfect score: 1437

Sequence: 1 ttcaaaactcataacgtgat.....tttatttgagtaaaaaaa 1437

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1437	100.0	1437	1 X02830	W09905287 Seq ID 5
2	35.8	2.5	480	1 Q81474	HMG2 promoter II
3	35.8	2.5	501	1 N50027	Sequence encoding
4	35.6	2.5	501	1 N50029	Sequence encoding
5	34.8	2.4	4951	1 T05848	Polycistronic cps
6	34.6	2.4	4093	1 Q49264	ced-4. Long-distal
7	34.4	2.4	1755	1 T67236	Pasteurella haemol
8	34	2.4	7778	1 V74359	Staphylococcus aur
9	33.8	2.4	11764	1 X12885	Enterococcus faeca
10	33.4	2.3	501	1 N50028	Sequence encoding
11	33.4	2.3	110000	1 V21209_13	Continuation (14 o
12	33.2	2.3	1078	1 X20337	Borrelia burgdorfe
13	33	2.3	26764	1 T71696	Human deoxycytidyl
14	32.8	2.3	2079	1 X07338	Streptococcus pneu
15	32.8	2.3	2339	1 T98581	DNA encoding a S
16	32.8	2.3	2511	1 X07337	Streptococcus pneu
17	32.8	2.3	7913	1 V45588	Plasmid pUC-CONTROL
18	32.8	2.3	12835	1 V52274	Streptococcus pneu
19	32.8	2.3	12845	1 V84274	Streptococcus pneu
20	32.8	2.3	19124	1 T72882	Plasmodium var-7 g
21	32.6	2.3	501	1 N50024	Sequence encoding
22	32.6	2.3	1579	1 X27359	Human secreted pro
23	32.6	2.3	6146	1 X03045	Human IL-1ra BAC c
24	32.6	2.3	6727	1 X02993	Human IL-1ra BAC c
25	32.4	2.3	11225	1 N91595	Sequence of fowipo
26	32.4	2.3	110000	1 X20448_03	Continuation (4 of
27	32.2	2.2	2941	1 O52490	Helminth aminopept
28	32.2	2.2	3358	1 O52501	Helminth aminopept
29	32.2	2.2	5189	1 V65701	DNA fragment of a
30	32.2	2.2	9834	1 V74348	Staphylococcus aur
31	32.2	2.2	110000	1 V21209_14	Continuation (15 o
32	32	2.2	729	1 N91692	Part of sequence o
33	32	2.2	3974	1 T43360	Cotton Folate2-82A

34 32 2.2 4721 1 V74545 Staphylococcus aur
35 32 2.2 5518 1 T73870 Cotton fibre promo
36 32 2.2 5547 1 T73865 Cotton fibre promo
37 31.8 2.2 521 1 V87552 EST clone DM324. N
38 31.8 2.2 1312 1 V27047 xenopus neurogenin
39 31.8 2.2 2831 1 Q87043 Murine U2 afbp-rs
40 31.8 2.2 5336 1 T90522 Arabidopsis male f
41 31.6 2.2 713 1 X00702 Human secreted pro
42 31.6 2.2 717 1 X00626 Human secreted pro
43 31.6 2.2 6412 1 Q79622 Flecked Morning GI
44 31.6 2.2 8011 1 V38336 Manic-depressive I
45 31.6 2.2 8065 1 V38335 Manic-depressive I

ALIGNMENTS

RESULT 1
X02830
ID X02830 standard; CDNA; 1437 BP.
AC X02830;
DT 14-MAY-1999 (first entry)
DE W09905287 Seq ID 5.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Perilla frutescens.
FH Key Location/Qualifiers
FT CDS 294..1301
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
FT PN W09905287-A1.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 23-JUL-1997; JP-200571.
PA (SUNR) SONTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR P-PSDB; W92951.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
PS Disclosure: Page 64-67; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.
CC Sequence 1437 BP; 375 A; 277 C; 325 G; 460 T;
SQ

Query Match 100.0%; Score 1437; DB 1; Length 1437;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttcaaaactcataacgtgatgagctaatgacatcttctccttcaagtcacagt 60
DB 1 TTCAAAACTCATACGTGATGAGTATGTCACATCTTCCCTTCAAGTCTACAGTG 60
QY 61 tcatctaccagcatcatcatgatactctcttataatgaggagaatggagtaacaag 120
DB 61 TCATCTACCAGCATCATCATGATCAATCTCTTTTAAATGAGGAGAAATGAGTAAGAAGG 120
QY 121 agtgggttttttactcagcttcaacctacgtactactactactactactactactcaag 180
DB 121 AGTGGGTTTTTGTACTCAGCTTCAACCTACGTAAGTACTACTACTACTCAACTCTCAG 180
QY 181 agaatgaataataataatagggcgatagatcttttagatatgtagggtgagcttgcga 240
DB 181 AGAATGAATATATATATAATAGGGCGATAGATCTTTTGTAGATATGTAGGTGTAGCTGCA 240
QY 241 ggtgggttaatttcgggtgtgggaaaaataataataataataataataataataatagc 300
DB 241 GGTGGTAAATTAATTTCCGGTGTGGGAAAAATAATAATAATAATAATAATATATAGCGATGAGCA 300

[illegible]


```
Query Match      2.4%; Score 33.8; DB 1; Length 11764;
Best Local Similarity 54.4%; Pred. NO. 15;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 171 aactctcaagagaataataataataatgagcgatagatcttcttagatagataggt 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 AAGATTAAATGATTAATAACACAGCAAAATATAAGATTTTGTCTTAACTCTTTTGT 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 231 gtgcctcgagggggtgtaattattccgggtggtgggaaataaataaataaataata 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 TTCAATAGGAAGTCTTACTGTCTTTTTCAGAAATGAAATGCAATACGACAAAAATTTA 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 291 gcgat 295
    |||
Db 193 TCTAT 197
    |||

RESULT 10
N50028/c
ID N50028 standard; DNA; 501 BP.
AC N50028;
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN-445.
KW Antiviral; cell growth regulator; immune system regulator;
OS Antiproliferative; ss.
FH Key Location/Qualifiers
FT cds 1..501 /*tag= a
FT FT
PN EP-163993-A.
PD 11-DEC-1985.
PF 17-MAY-1985. 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
DR WPI: 85-311944/50.
DR P-PSDB: P50027.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 21, page 37; 7ipp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 114 A; 31 C; 68 G; 79 T;

Query Match      2.3%; Score 33.4; DB 1; Length 501;
Best Local Similarity 28.3%; Pred. NO. 4.1;
Matches 60; Conservative 49; Mismatches 103; Indels 0; Gaps 0;

Qy 796 ttctgtcataacaggctctattttgttcacactcagcactgttggtcagaataactg 855
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 TTCANCCNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 856 tgaagaggaagatcgattgaagtagtagcgaatgcggtgattgttgatgcataca 915
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 TADATNGTVARNCGRCRCYCTYTTTGRAAYTGTGNCCTGYTTDATYTCVTCNGGD 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 916 gtcaacagagatctgctgcgagaggaatgccctctcaatcaatgactgctactg 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ATRTCRAARTTCATGKRCYCTTVARCAATAYTCVARCHNCCRTTVARTGCCAVRV 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 976 gactttggcagcctatgtttgtttgactg 1007
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 ARATTTTGCAATGCAATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 28
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
V21209_13/c
Continuation (14 of 17) of V21209 from base 1300001 (Methanococcus jannaschii circula
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match      2.3%; Score 33.4; DB 1; Length 110000;
Best Local Similarity 52.5%; Pred. NO. 55;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 151 gtacgtactactactgactcaactcaagagaagaataataataataataatgagcgatag 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33193 GTACACGTATCAACGATGAATTAATAAGAGAAATTTAAATAGCTAAGACGCTAAAC 33134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 211 atctttgtagatatgtagggtgtagctgcaggtgaggttaatttaatttcgggtgggaaaa 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33133 AAATTTGTAAATTCCTTCAACAAGACAGAGAGAGTAAATTTTATCATCCTGCTTATACA 33074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 271 taataataataataataat 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33073 GAGGTTGATGAAGTAATAT 33055
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
X20337
ID X20337 standard; DNA; 1078 BP.
AC X20337;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #90.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; GS.
OS Borrelia burgdorferi.
PN W09858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 107; 1129pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
```


US5622851-A.
22-APR-1997.
PD
PF 10-JAN-1995; 370975.
PR 10-JAN-1995; US-370975.
PA (HEAL-) HEALTH RES INC.
FI Mailey F., Mailey GR., Weiner KXB;
DR WPI; 97-244391/22.
DR P-PSDB; W18205.
DR DNA encoding human deoxycytidylate deaminase - for production of
PT recombinant deaminase
PT Claim 3; Column 55-78; 58pp; English.
PS

DR P-PSDB: W97801.
PT New SecA polypeptides and polynucleotides - useful as diagnostic
PT reagents and for prevention and treatment of Streptococcus
PT pneumoniae infections, especially meningitis
PT
PS Claim 27; Page 28-29; 34pp; English.
CC This DNA sequence encodes novel SecA polypeptide (see W97801) of
CC Streptococcus pneumoniae 010093 (NCIMB 40794). A polynucleotide
CC having this DNA sequence can be isolated from a library of clones
CC of chromosomal DNA of *S. pneumoniae*. A vector comprising the
CC polynucleotide, host cells and a method for producing a SecA
CC polypeptide or fragment are claimed. SecA polypeptides and
CC polynucleotides are useful for diagnosing diseases related to over
CC or underexpression of SecA protein by identifying mutations in the
CC SecA gene, or determining SecA polypeptide or mRNA expression
CC levels due to an infection of an organism with the SecA gene
CC (claimed). They can diagnose the stage and type of infection.
CC SecA PCR probes are useful for diagnosing diseases, and can
CC characterise the response of the infectious organism to drugs.
CC SecA polypeptides and polynucleotides are also useful for screen
CC for antagonists, agonists and drugs against infectious
CC microorganisms. SecA polypeptides and antagonists are useful for
CC treating conditions associated with abnormal SecA protein levels
CC (claimed), and SecA agonists, antagonists and drugs are
CC bacteriostatic and bacteriocidal compounds which can be used in
CC treatment to enhance (agonist) or block (antagonist or antisense
CC sequence) SecA activity, thereby treating microbial diseases.
CC

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 13:21:41 ; Search time 9137.45 Seconds
(without alignments)
280.670 Million cell updates/sec

Title: US-09-147-955-9
Perfect score: 1437
Sequence: 1 ttcaaaactcataacgtgat.....tttatttgagtaaaaaaaaa 1437

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: em_fun.*
17: em_hum1.*
18: em_hum2.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_sy.*
29: em_un.*
30: em_vi.*
31: gb_htg1.*
32: gb_htg2.*
33: gb_in1.*
34: gb_in2.*
35: em_ba1.*
36: em_ba2.*
37: em_hum3.*
38: em_hum4.*
39: gb_pr4.*
40: gb_htg3.*
41: gb_htg4.*
42: gb_htg5.*
43: gb_htg6.*

44: gb_htg7.*
45: em_htg1.*
46: em_htg2.*
47: em_htg3.*
48: em_hum5.*
49: gb_pl3.*
50: gb_pr5.*
51: gb_htg8.*
52: gb_htg9.*
53: gb_htg10.*
54: gb_htg11.*
55: gb_htg12.*
56: gb_htg13.*
57: gb_htg14.*
58: gb_in3.*
59: gb_htg15.*
60: gb_htg16.*
61: gb_htg17.*
62: em_htg4.*
63: em_htg5.*
64: em_htg6.*
65: em_htg7.*
66: em_hum6.*
67: gb_htg18.*
68: gb_htg19.*
69: gb_htg20.*
70: gb_htg21.*
71: gb_htg22.*
72: gb_htg23.*
73: gb_htg24.*
74: gb_htg25.*
75: gb_htg26.*
76: gb_htg27.*
77: gb_htg28.*
78: gb_htg29.*
79: gb_htg30.*
80: gb_htg31.*
81: gb_v11.*
82: gb_v12.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	92.8	6.5	4005	7	SPCC320
2	45	3.1	7218	5	I66494
3	43.8	3.0	65870	12	AC005807
c 4	43.6	3.0	7218	5	I66494
5	43	3.0	96958	39	AC006350
6	42.4	3.0	98630	49	AC006535
7	42.2	2.9	1857	12	MUSMHTL1
8	42.2	2.9	5187	12	MUSMHLA
c 9	41.6	2.9	148241	41	AC009857
c 10	41.6	2.9	167429	70	AC010952
c 11	41.6	2.9	196085	32	CNS01RGW
12	41.2	2.9	200337	61	AC021186
13	41	2.9	188461	42	AC010949
c 14	41	2.9	189910	40	CNS01RG0
c 15	40.8	2.8	22696	42	AC012895
c 16	40.8	2.8	156492	54	AC008460
17	40.8	2.8	218596	60	AC021781
18	40.8	2.8	327311	34	AE003500
c 19	40.6	2.8	102263	32	AL139152
c 20	40.6	2.8	167101	60	AC015595
21	40.4	2.8	84412	10	HS973M2
c 22	40.4	2.8	101154	8	ATFLP2
c 23	40.2	2.8	158546	54	AC022876
c 24	40.2	2.8	181671	31	AP001024

```

25 40.2 2.8 200038 43 AC018423 Homo sapi
26 40 2.8 58834 51 AC022723 Homo sapi
27 40 2.8 163048 54 AC0122101 Homo sapi
28 40 2.8 176008 40 AL353113 Homo sapi
29 39.8 2.8 456 9 HSN0V1
30 39.8 2.8 59654 57 AC025727
31 39.8 2.8 110000 31 PFMAL13p2_2
32 39.6 2.8 85992 7 AB009052 Arabidops
33 39.4 2.7 76175 49 AC009273
34 39.4 2.7 97189 10 HS479J7
35 39.2 2.7 114534 31 AP000633
36 39.2 2.7 152981 31 AP001193
37 39.2 2.7 159164 31 AP001140
38 39.2 2.7 172078 40 AL160276 Homo sapi
39 39.2 2.7 183461 39 AC006121
40 39 2.7 159553 44 AC009990
41 39 2.7 203826 74 AC021130
42 38.8 2.7 8558 33 CEU95074
43 38.8 2.7 29350 33 CEC14B1
44 38.8 2.7 68794 72 AC034112
45 38.8 2.7 130514 60 AC016441

```

ALIGNMENTS

RESULT 1

```

SPCC320/c 40005 bp DNA PLN 06-APR-1999
LOCUS S.pombe chromosome III cosmid c320.
DEFINITION AL022245
ACCESSION AL022245
VERSION 2
KEYWORDS beta-synthase; bio2, biotin synthase; EF-hand calcium-binding
domain; ferredoxin; fungal Zn(2)-cys(6) binuclear cluster
domain signature; protein kinase; putative transporter; ribosome
subunit biogenesis; rna binding protein; serine/threonine
dehydratase; signal recognition particle protein subunit;
transcriptional regulator; trna- Glu.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Archiasscomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes;
1 (bases 1 to 40005)
Wood,V., Rajandream,M.A., Barrell,B.G., Wedler,H., Wambutt,R. and
Wedler,E.
Direct Submission
Submitted (24-MAR-1998) European Schizosaccharomyces genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and AGOWA GmbH, Glienicke Weg 185, D-12489 Berlin, Germany
On Apr 14, 1999 this sequence version replaced gi:2995362.
Notes:
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites. CDS are numbered using the following
system eg SPCC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2
(cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Details of yeast
sequencing at the Sanger Centre are available on the World Wide
Web.
(URL, http://www.sanger.ac.uk/Projects/S.pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was

```

```

FEATURES
source
1..40005
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="III"
/clone="cosmid c320"
/map="III"
misc_feature
1..7687
/note="nominal overlap with cosmid SPCC330 S. pombe
chromosome 3"
complement(709..1680)
/gene="spcc320.14"
complement(709..1680)
/gene="spcc320.14"
/note="SPCC320.14, len:322, SIMILARITY: Saccharomyces
cerevisiae, VKL218C, YKV8_YEAST, P36007, hypot helical
protein with strong similarity to E.coli and H.influenzae
threonine dehydratases, (396aa), fasta scores, Opt:14 64,
E():10, (66.4% identity in 318 aa overlap)"
/codon_start=1
/product="putative serine/threonine dehydratase"
/protein_id="CAA18316.1"
/db_xref="GI:2995376"
/db_xref="SPTREMBL:059791"
/translaton="MSDNLYLPTVDVVASASERIKKPAKTPVLTSSTVKEFVAEVF
PKCFENFGKAFKRGALNALSNLNEAQRKAGVLFSSNHAQIALAKILGIPAKI
IMPLDAPKVAATGKGGVIMYDRTKODREKMAKEISERGLTIIPYDHPHYLAG
QTAKEALFEVGLPDALFVCLGGGLSGSALAARHAFNPCEVGVPEAGNQGQOS
FRKSVIHDPKTIADGAQTHLGNVTFSTIKERKVDILTVSDEELDKLFYAARM
KIVPEPTGCLSFSAARAMEKRLKNGIIGISGNVDIERVAHFLSO"
complement(1123..1518)
/gene="spcc320.14"
misc_feature
1..40005
/note="Pfam match to entry S_T_dehydratase ,
beta-synthase"
2454..2525
/gene="trna Glu"
/product="trna Glu"
2454..2525
/gene="trna Glu"
3062..4216
/gene="spcc330.16"
3062..4216
/gene="spcc330.16"
/note="SPCC330.13c, len:384, SIMILARITY: Xenopus laevis,
Q91820, P46EG265, The xenopus protein kinase p32
associates with the centrosome in a cell cycle-dependent
manner, binds to the spindle microtubules and is involved
in bipolar mitotic spindle assembly., fasta scores,
opt:1185, E():0, (48.2% identity in 380 aa overlap), also
similar eg. to p97477, serine/threonine protein kinase,
(395aa), fasta scores, opt:1157, E():0, (54.9% identity in
317 aa overlap)"
/codon_start=1
/product="ser/thr protein kinase"
/protein_id="CAA18315.1"
/db_xref="GI:2995375"
/db_xref="SPTREMBL:059790"
/translaton="MVLPOVNNNAPTSGRTCKPTSAFFVEIMSDSKLADSLCLSV
STPSTTANFGQQLRLAYSNQVNVNLANGKENKRTSNFKNSIRKEETPIAGV
PSSGQWRFPHGMFEIGKRGKRGVRLAKKKTGFVALKTLHKSELVQKVG
KQVREIETSNLHNNILRYGHFKRIYLLIEFAGRGYLOHLRRKRRFSEVKA
SKYITFQNALSYLHKHVIHRDKPENILLIDGELKLSDFGNSVHAPSRRRTLCG
TLDYLPPEMVEGKEHTEKVDLSGLVLYTEFLVGAPPFEDMSGHSATYKRIKVDLKI
PSFVPPDARDLISRLQHNPEKRMSLSEQVMRPWTVKYKDSWTRKSSSS"
3413..4168
/gene="spcc330.16"
misc_feature

```

sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project. Cosmid c320 is overlapped by cosmid c330 at the 5' end and c1235 at the 3' end.

```

/misc_feature
/note="Pfam match to entry pkinase PF00069, Eukaryotic
prot ein kinase domain"
3431. .3502
/gene="spcc330.16"
/note="P500107 Protein kinases ATP-binding region
signature"
3770. .3808
/gene="spcc330.16"
/note="P500108 Serine/Threonine protein kinases
active-site signature"
complement(4707. .5330)
/gene="spcc320.12"
complement(join(4707. .4714,4781. .5330))
/gene="spcc320.12"
/note="SPCC320.12, len:184, SIMILARITY: Saccharomyces
cerevisiae, YNR020C, YN8A_YEAST, hypothetical protein,
(270aa), fast a scores, opt:592, E():0, (47.5% identity in
183 aa overlap)"
/codon_start=1
/product="hypothetical 21.5 kd protein"
/protein_id="CAAL8314.1"
/db_xref="GI:2995374"
/db_xref="SPTREMBL:O59789"
/translation="MSEDPQSKERKNCERVKRALMSQSPVILFLKALDRINCNI
EKDISCPDAQSTGGTIPGKGVLCENKRLTKKMAENTIAHEMIHFDHREVDWN
NLRHQACSEIRASSGECRWKELRFGNIKTRKHHCQVKRRATISVQGNPKSK
EQEAIVERVFNSCFNDRPPEKTY"
complement(4715. .4738)
/gene="spcc320.12"
/note="ctaaactattcattcttcgtctag, splice branch and
acceptor"
complement(4775. .4780)
/gene="spcc320.12"
/note="gtacgt, splice donor sequence"
complement(5055. .5084)
/gene="spcc320.12"
/note="P500142 Neutral zinc metalloproteinases,
zinc-binding region signature"
5867. .6724
/gene="spcc320.11c"
join(5867. 5922,6076. .6162,6325. .6724)
/gene="spcc320.11c"
/note="SPCC320.11c, len:179, SIMILARITY: Saccharomyces
cerevisiae, YPL211W, O08962, required for efficient 60S
ribosome subunit bio genesis, (181aa), fasta scores,
opt:892, E():0, (70.6% iden tity in 180 aa overlap)"
/codon_start=1
/product="putative ribosome biogenesis protein"
/protein_id="CAAL8313.1"
/db_xref="GI:2995373"
/db_xref="SPTREMBL:O59788"
/translation="MRPLTHEETKTFEKLQAYIGKNITHLIDRDPDHPGFLQKQRV
YVYSERAMKATSVARQNLMSGICFGKFTNKRFLHITALDYIAQYARYKIWKVSN
GEMPLYGNHVLKHAHVGRTIDTTPQHGVIYSNMDTPLGFGVTARSTLELRLEPTA
IVAFHQADVCEYLDRDETLF"
5923. .5928
/gene="spcc320.11c"
/note="gtaagt, splice donor sequence"
6059. .6075
/gene="spcc320.11c"
/note="ctgactagtaaaactag, splice branch and acceptor"
6163. .6168
/gene="spcc320.11c"
/note="gtaaga, splice donor sequence"
6310. .6324
/gene="spcc320.11c"
/note="ttaacatactatag, splice branch and acceptor"
complement(7697. .9615)
/gene="spcc320.10"
complement(join(7697. .8738,8785. .9104,9156. .9218,
9355. .9615))
/gene="spcc320.10"
/note="SPCC320.10, len:560, SIMILARITY:Homo sapiens,

```

```

SR72_HUMAN, signal recognition particle 72 kd protein,
(670 aa), fasta scores: opt: 354, E():1.4e-14, (24.6%
identity in 574 aa)"
/codon_start=1
/label="SPCC320.10"
/product="putative signal recognition particle protein"
/protein_id="CAAL8312.1"
/db_xref="GI:2995372"
/db_xref="SWISS-PROT:O59787"

```

```

Query Match      6.5%; Score 92.8; DB 7; Length 40005;
Best Local Similarity 47.4%; Pred. No. 5.4e-13;
Matches 314; Conservative 0; Mismatches 342; Indels 6; Gaps 1;

```

```

QY 537 tatgtctctcctggggtctgtgacgtccacggcaattatctgttaataaagcattcgag 596
Db 13681 TATATCATCTTTGGCTTTTGGACGTTGAAGAAATATTTCCTGTATAGTACGACAA 13622
QY 597 ttgacatcgattacgagtgtagcatactgattgttggacaatcgtgtggccatcac 656
Db 13621 TACACAAACATGCTTAGTGCAAGTTTACTCGATTCTTTGGGCTACTGTGCAGTGTGAAT 13562
QY 657 tttaaatgaggttctctag-----gcacaaataactctgtataacagtttctcggtgt 710
Db 13561 CTTTCTCTTATCTTTTAAAGGTTTCGCTATCAGCTGGTCTCAAAATCTTGGTGTGCGGCT 13502
QY 711 gctatttgttagaggagcctcctcctcgtctcttcttccgactcaggggtcactgctgt 770
Db 13501 TGCATTGGTGGTTGGTTTGGCTGTGCTAGTGATGTATTCTCGTGGTGATTATCT 13442
QY 771 gtttgcgaatcctcttgggtgatttctgtcataacagagctctatttggcacactc 830
Db 13441 GCTGTCAACCCCTGGTTGGGTGGTGGTTACATGATTTATCGCGCTACGTGTATGGTGTG 13382
QY 831 agcactgttggtcagggaatactgcgtgaaaggaaagatcgtattgaagtagtagcaaty 890
Db 13381 TCCAAATACTTTGGAGGAATACTTTGCTTCCAAATGCGCTTTACGTAGTCAATCGGTCAA 13322
QY 891 atcgtgtatttggtagtctcctcagtcgacacgagagattactgctgagaggagaatgcc 950
Db 13321 TTAAGTCTCTACGGTCTCTATTATCAGCATTTATTCAAACCTTTATTTTGACCGTCAACAT 13262
QY 951 ctctcatcaatgagtggtctactggacttlttggcagcctatgttgttatgactgtctc 1010
Db 13261 TTGTATACCTTCATTTGGACTAGTGAATGGTGGTGTACTTAGCTGGTTTATTTCTCGTC 13202
QY 1011 agcttctctctgcacactcaaccccttctctcgaagatgagtggtgcgtgatttctc 1070
Db 13201 ATGTTCTTGCCTTATAGTTTGGCTCCTATCTCTTTAGAATGCTTCTGCGACTTTTAT 13142
QY 1071 aatcttccatgcttaccatctgatgatgtgggctgttgcaattaggacattcatatacaac 1130
Db 13141 AATATCTCTTTATTGACTTCCGACTTTTGGAGCTTGGTGTATTGTGTTATTCATGTTTGA 13082
QY 1131 caggaggttgattggttatactatttggccttttggctctctgttcttattggaatattcata 1190
Db 13081 TATCATGTGCTACTGGCTGTACCCCATTCGCTTCTTAATATATATGTTGGGTTTGTGTA 13022
QY 1191 ta 1192
Db 13021 TA 13020

```

```

RESULT 2
I66494
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

```

[illegible]

```

Qy 202 gggcagatagatttttagatattagtgtagctgcagtggtgtaatttaattccgggt 261
Db 1328 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1269
Qy 262 gtgggaataataataataatagcagcagcagcagcagcagcagcagcagcagcagcag 321
Db 1268 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1209
Qy 322 gacagataggggagcgaagacattgctgggttggttggttggtggcag 371
Db 1208 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1159

RESULT 5
AC006350 LOCUS AC006350 Homo sapiens PAC clone RP5-998H4 from 7q21.1-q21.2, complete
DEFINITION AC006350 Homo sapiens PAC clone RP5-998H4 from 7q21.1-q21.2, complete
ACCESSION AC006350
VERSION AC006350.2 GI:4508140
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 96958)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 96958)
AUTHORS Madsen,C. and Spalding,L.
TITLE The sequence of Homo sapiens PAC clone RP5-998H4
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 96958)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 96958)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 96958)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 96958)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 24, 1999 this sequence version replaced gi:4139372.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_DJ0998H04
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RP5-998H4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTB-22J17, 200 bp overlap. Actual start of this clone is at base position 1 of RP5-998H4; actual end is at 27949 of CTB-22J17.

FEATURES

source

1..96958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21.1-q21.2"
/clone="RP5-998H4"
/clone_lib="RP5-998H4"
799..881
/rpt_family="L1"
882..1242
/rpt_family="MaLR"
1243..2444
/rpt_family="MaLR"
2445..2753
/rpt_family="Alu"
2754..3185
/rpt_family="MaLR"
3186..3547
/rpt_family="MaLR"
3548..3652
/rpt_family="L1"
4058..4356
/rpt_family="Alu"
4708..4991
/rpt_family="Alu"
4992..5017
/rpt_family="(TAA)n"
5741..5810
/rpt_family="L2"
5842..6141
/rpt_family="Alu"
6313..8075
/rpt_family="L1"
8228..8268
/rpt_family="(CA)n"
8772..8865
/rpt_family="MER21_g"
8900..9034
/rpt_family="MIR"
9742..9866
/rpt_family="MIR"
10829..10874

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

Query Match 3.0%; Score 43; DB 39; Length 96958;
Best Local Similarity 50.7%; Pred. No. 1.8;
Matches 103; Conservative 0; Mismatches 100; Indels 0

QY	1234	agaatggaacttgatcatgaatatagctctcttgaggatcaaatgatcacacccaagaa	1293
Db	69232	ACAGTAGACAATAGGTCATCAATAAAGAAACCATGCCAAGTAGTAATGAAGRAGACNA	69291
QY	1294	aaccatagctagcttgcccaaatctcttccatcaacagttttaataatctcgagggg	1353
Db	69292	TAGCTTTCTTTCTTTCAAAAATAGTGTTCATCAGCATTTCTATGACATCTTCAAGGGT	69351
QY	1354	gagagagatcgagatcaactaattaatggacgctctattatattatgttgaggtttt	1413
Db	69352	TATAGATGAAGGTAGACCAATAGCTAGGTGGAGTGTGTGCTTAAATGTTTTATTGT	69411
QY	1414	tttatttatttgagtaaaaaaa	1436
Db	69412	TACAATTAATTTTCAAGTTAAAA	69434
RESULT	6		
LOCUS	AC006535	AC006535 98630 bp DNA PLN 15-FEB-2000	
DEFINITION		Genomic sequence for Arabidopsis thaliana BAC T24P13 from chromosome I, complete sequence.	
ACCESSION	AC006535	AC006535	
VERSION	AC006535.7	GI:6978061	
KEYWORDS	HTG.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
AUTHORS		Johnson-Hopson, C., Dunn, P., Brooks, S., Buehler, E., Chao, Q., Khan, S., Kim, C., Shinn, P., Altati, H., Bel, Q., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howling, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharshy, N., Nguyen, M., Paim, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Federspiel, N.A., Theologis, A. and Ecker, J.R.	
TITLE		Genomic sequence for Arabidopsis thaliana BAC T24P13 from chromosome I	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 98630)	
AUTHORS		Ecker, J.R.	
TITLE		Direct Submission	
JOURNAL		Submitted (07-FEB-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	
REFERENCE		3 (bases 1 to 98630)	
AUTHORS		Ecker, J.R.	
TITLE		Direct Submission	
JOURNAL		Submitted (12-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	
REFERENCE		4 (bases 1 to 98630)	
AUTHORS		Ecker, J.R.	
TITLE		Direct Submission	
JOURNAL		Submitted (15-FEB-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA	
COMMENT		On Feb 15, 2000 this sequence version replaced gi:4803878.	
FEATURES		Location/Qualifiers	
source		1. 98630	
		/organism="Arabidopsis thaliana"	
		/db_xref="taxon:3702"	
		/chromosome="1"	
		/clone="T24P13"	
BASE COUNT	32079 a 17075 c 16743 g 32733 t		
ORIGIN			
Query Match	3.0%	Score 42.4;	DB 49; Length 98630;


```

Best Local Similarity 45.13; Pred. No. 2.5; Mismatches 166; Indels 0; Gaps 0;
Matches 142; Conservative 0;

QY 1075 titccatctacatcctgatactgtggctgttgcgaattaggacattacacaaccagg 1134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83138 TTTTCCATATTAATATCTTTAGTTAGTTGTTTAAATGAAATTTGTTTAAATATT 83197
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1135 aggttgattggtataactattgtgcctttgtctcctgttggtagaatattacatatata 1194
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83198 TACTATATTTTAAATGGAATTCCTTTTCTTAAATATATATAGTTGGAATTTATATATA 83257
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1195 caaaacagagaagagctcctaacaatcagagagcccttgagaatggaacttgggacatg 1254
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83258 ATATTAATGACAAATTTTGTAAATCTAGTAAATGATGAACAAGAAATTAAGTTGGTT 83317
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1255 aaatagctccttgagatgaagaatgacacacacaagaacacatagctgttgcacca 1314
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83318 GAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 83377
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1315 caatcttttccatcacaggttttaataattcgtgaggggagagatcgagatcactaat 1374
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83378 AGGTATATTTTCTTAAGTATGTTCTTCTTAAATATATATAGGGATAGAGATCTTATCAT 83437
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1375 taatggac 1382
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83438 TTTTTCAC 83445
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS MUSMHTLAI 1857 bp DNA ROD 27-APR-1993
DEFINITION Mouse MHC class I thymus leukemia antigen (Tla) gene, exons 1, 2
and 3
ACCESSION M11742.1 GI:199651
VERSION M11742.1
KEYWORDS Alu-like repeat; cell surface antigen; cell surface differentiation
antigen; cell surface glycoprotein; class I gene; integral membrane
protein; leukemia antigen; major histocompatibility complex.
SEGMENT 1 of 2
SOURCE Mouse (C57BL/6) leukemia ERLD DNA, clone C25.1.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1857)
AUTHORS Obata,Y., Chen,Y.-T., Stockert,E. and Old,L.J.
TITLE Structural analysis of TL genes of the mouse
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 82, 5475-5479 (1985)
MEDLINE 85270545
FEATURES
Location/Qualifiers
source 1..1857
/organism="Mus musculus"
/db_xref="taxon:10090"
<1..44
/feature="thymus leukemia antigen"
/number=1
45..223
/note="Tla intron A"
224..493
/number=2
494..640
/note="Tla intron B"
641..916
/number=3
917..>1857
/note="Tla intron C"
BASE COUNT 494 a 422 c 429 g 512 t
ORIGIN BamHI site; chromosome 17.

Query Match 2.9%; Score 42.2; DB 12; Length 1857;
Best Local Similarity 55.98; Pred. No. 3.5;
Matches 80; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

```

QY 1060 ctgcatcttttccaatcttccatgcttcatctgatatgtggctgttgcgaattaggacat 1119
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1407 CTCCAAATATTAGCTTTCCCTGATTTTCATTTATGTCATGTCAGTAACACAGGTACAT 1466
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1120 tcatacaacaacagaggttgattggttatactattgaccttttgccttttgcctgtgttg 1179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1467 TTTCTAAACAAGTATACGTTTTTTTGTCTTTCTTTGTTCTTTTGTGTTGTTGTTGCC 1526
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1180 gaattatcatatatacaaaaaa 1202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1527 TAAATTACCTCTATGAGCAAAA 1549
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
LOCUS MUSMHLA 5187 bp DNA ROD 27-APR-1993
DEFINITION Mouse MHC class I T3-d gene (H-2-d haplotype).
ACCESSION M75875
VERSION M75875.1 GI:199554
KEYWORDS beta-2-microglobulin associated protein; cell surface glycoprotein;
class I antigen; class I gene; integral membrane glycoprotein;
major histocompatibility complex; thymus leukemia antigen.
SOURCE Mus musculus (strain BALB/c, sub-species domesticus) (library:
cosmid) sperm DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5187)
AUTHORS Mashimo,H., Chorney,M.J., Pontarotti,P.A., Fisher,D.A., Hood,L. and
Nathenson,S.G.
TITLE Nucleotide sequence of the BALB/c H-2T region gene T3
JOURNAL Immunogenetics 36, 326-332 (1992)
MEDLINE 92355117
FEATURES
Location/Qualifiers
source 1..5187
/organism="Mus musculus"
/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/cell_line="spermatzoa"
/germline
/nucleotide="H-2-d"
/tissue_type="sperm"
/tissue_lib="cosmid"
/map="Chromosome 17 H-2T r"
join(222..300,480..749,898..1173,2811..3086,3219..3332,
3628..3767)
/gene="MHC class I T3-d"
222..300
/gene="MHC class I T3-d"
/number=1
join(222..300,480..749,898..1173,2811..3086,3219..3332,
3628..3767)
/gene="MHC class I T3-d"
/codon_start=1
/product="MHC thymus leukemia antigen"
/protein_id="AAA39658.1"
/db_xref="GI:199555"
/translation="MRMGTPVPGTLLILAAAGQGTQTCGSHSLRYPTALSRPAIS
EPWYIAGYLDQTFQVFNSSGATYKLSAPWQEGPEYWARETEIVTNAOFFRE
NIQTMIDYNLSONGSHITQVMYGCVEFFGSLFRAYEHOHYDQPDYIALNEDLKTWT
AADTAAETRSKWEQAGTYELRRTYLEGPKDLSLLRYLNRKKTQCTDCKPTKHTVTH
PRPEGYVTLRCWALRYFPADITLTMQINGEELIOTDELVETEPAGDGFQKAAVVVP
LCKEQKTYCHVYHGLFLEPILTLRWEPPQISMNRTVYRALLGAMILLGFNGSVMMWM
RKNNGGNDONTIAIYQNEREHLSDPRAEALGVGAKMLDPSAPPLVS"
480..749
/gene="MHC class I T3-d"
/number=2
898..1173
/gene="MHC class I T3-d"
/number=3
2811..3086
/gene="MHC class I T3-d"

```

```

http://ftp.genome.washington.edu/rw/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

JOURNAL OF
REFERENCE
2 (bases 1 to 167429)

Lehoczeky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 2000 this sequence version replaced gi:6249728.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L2836
Center clone name: 4_N_16

* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1114: contig of 1114 bp in length
* 1115 1214: gap of 100 bp
* 1215 2323: contig of 1109 bp in length
* 2324 2423: gap of 100 bp
* 2424 3499: contig of 1076 bp in length
* 3500 3599: gap of 100 bp
* 3600 4691: contig of 1092 bp in length
* 4692 4791: gap of 100 bp
* 4792 6613: contig of 1822 bp in length
* 6614 6713: gap of 100 bp
* 6714 8003: contig of 1290 bp in length
* 8004 8103: gap of 100 bp
* 8104 9521: contig of 1418 bp in length
* 9522 9621: gap of 100 bp
* 9622 10894: contig of 1273 bp in length
* 10895 10994: gap of 100 bp
* 10995 12113: contig of 1119 bp in length
* 12114 12213: gap of 100 bp
* 12214 14077: contig of 1864 bp in length
* 14078 14177: gap of 100 bp
* 14178 15552: contig of 1375 bp in length
* 15553 15652: gap of 100 bp
* 15653 16845: contig of 1193 bp in length
* 16846 16945: gap of 100 bp
* 16946 18701: contig of 1756 bp in length
* 18702 18801: gap of 100 bp
* 18802 20280: contig of 1479 bp in length
* 20281 20380: gap of 100 bp
* 20381 22043: contig of 1663 bp in length
* 22044 22143: gap of 100 bp
* 22144 23464: contig of 1321 bp in length
* 23465 23564: gap of 100 bp
* 23565 25268: contig of 1704 bp in length
* 25269 25368: gap of 100 bp
* 25369 26869: contig of 1501 bp in length
* 26870 26969: gap of 100 bp
* 26970 28961: contig of 1992 bp in length
* 28962 29061: gap of 100 bp
* 29062 31160: contig of 2099 bp in length
* 31161 31260: gap of 100 bp
* 31261 32989: contig of 1729 bp in length
* 32990 33089: gap of 100 bp

* 33090 35728: contig of 2639 bp in length
* 35729 35828: gap of 100 bp
* 35829 37664: contig of 1836 bp in length
* 37665 37764: gap of 100 bp
* 37765 39295: contig of 1531 bp in length
* 39296 39395: gap of 100 bp
* 39396 41939: contig of 2544 bp in length
* 41940 42039: gap of 100 bp
* 42040 44376: contig of 2337 bp in length
* 44377 44476: gap of 100 bp
* 44477 47182: contig of 2706 bp in length
* 47183 47282: gap of 100 bp
* 47283 50373: contig of 3091 bp in length
* 50374 50473: gap of 100 bp
* 50474 54496: contig of 4023 bp in length
* 54497 54596: gap of 100 bp
* 54597 57825: contig of 3229 bp in length
* 57826 57925: gap of 100 bp
* 57926 64181: contig of 6256 bp in length
* 64182 64281: gap of 100 bp
* 64282 70174: contig of 5893 bp in length
* 70175 70274: gap of 100 bp
* 70275 74993: contig of 4719 bp in length
* 74994 75093: gap of 100 bp
* 75094 80432: contig of 5339 bp in length
* 80433 80532: gap of 100 bp
* 80533 85565: contig of 5033 bp in length
* 85566 85665: gap of 100 bp
* 85666 95772: contig of 10107 bp in length
* 95773 95872: gap of 100 bp
* 95873 108321: contig of 12449 bp in length
* 108322 108421: gap of 100 bp
* 108422 118802: contig of 10381 bp in length
* 118803 118902: gap of 100 bp
* 118903 130159: contig of 11257 bp in length
* 130160 130259: gap of 100 bp
* 130260 146057: contig of 15798 bp in length
* 146058 146157: gap of 100 bp
* 146158 167429: contig of 21272 bp in length.

FEATURES Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-4N16"
/clone_lib="RPC1-11 Human Male BAC"
1. 1114
misc_feature
1215 2323
/note="assembly_fragment"
misc_feature
2424 3499
/note="assembly_fragment"
misc_feature
3600 4691
/note="assembly_fragment"
misc_feature
4792 6613
/note="assembly_fragment"
misc_feature
6714 8003
/note="assembly_fragment"
misc_feature
8104 9521
/note="assembly_fragment"
misc_feature
9622 10894
/note="assembly_fragment"
misc_feature
10995 12113
/note="assembly_fragment"
misc_feature
12214 14077
/note="assembly_fragment"
misc_feature
14178 15552
/note="assembly_fragment"
misc_feature
15653 16845
/note="assembly_fragment"
misc_feature
16946 18701
/note="assembly_fragment"
misc_feature
18802 20280
/note="assembly_fragment"

sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, phage, etc.	
Contig order : 5 4 3, 1000 N's separate segments	
Contig 5 : length 102436 bp	
Contig 4 : length 55923 bp	
Contig 3 : length 3720 bp	
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.	
* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.	
* 1 102436: contig of 102436 bp in length	
* 102437 103436: gap of 1000 bp	
* 102437 159365: contig of 55923 bp in length	
* 159366 160365: gap of 1000 bp	
* 160366 196085: contig of 35720 bp in length.	
* Location/Qualifiers	
1..196085	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="14"	
/clone_lib="RPC1-11"	
/clone="R-355C3"	
72179..72308	
/note="matching EMBL:G07705; RHdb:RH32619; dbSTS:STS32292; identified using the e-PCR software (G. Schuler)"	
95324..95487	
/note="matching EMBL:Z39869; RHdb:RH53631; dbSTS:STS17739; identified using the e-PCR software (G. Schuler)"	
62367 a 36418 c 36354 g 58939 t 2007 others	
BASE COUNT	
ORIGIN	
Query Match	
Best Local Similarity 2.9%; Score 41.6; DB 32; Length 196085;	
Matches 80; Conservative 0; Mismatches 64; Indels 0; Gaps 0;	
QY 1293 aaaccatagctgcttggccacacatcttccatcaacagctttaaataatcgtagggg 1352	
DB 129261 AAACATAGATATTTTCTTAATAACTGTGTAATAGGGTTTGTGTAGTAAG 129202	
QY 1353 ggagagagatcgagatctaatgaatgagctcttatcatagctggaggtttgtttt 1412	
DB 129201 CTAGTGTCAAAAATAATGAAACTAACTCTTGCTATAGTATATTTTCTTTT 129142	
QY 1413 attattattttgagtaaaaaaa 1436	
DB 129141 GTTCACTCAATGGGAAAAGGAAA 129118	
RESULT 12	
AC021186 200337 bp DNA HTG 19-MAR-2000	
LOCUS	
DEFINITION Homo sapiens chromosome UL clone RP11-744D14, WORKING DRAFT	
SEQUENCE, 7 unordered pieces.	
AC021186	
ACCESSION AC021186.2 GI:7105593	
VERSION	
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE human.	
ORGANISM Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 200337)	
REFERENCE	
AUTHORS Waterston,R.H.	
TITLE The sequence of Homo sapiens clone	
JOURNAL Unpublished	
COMMENT 2 (bases 1 to 200337)	
AUTHORS Waterston,R.H.	
20381..22043	
/note="assembly_fragment"	
22144..23464	
/note="assembly_fragment"	
23565..25268	
/note="assembly_fragment"	
25369..26869	
/note="assembly_fragment"	
26970..28961	
/note="assembly_fragment"	
29062..31160	
/note="assembly_fragment"	
31261..32989	
/note="assembly_fragment"	
33090..35728	
/note="assembly_fragment"	
35823..37664	
/note="assembly_fragment"	
37765..39295	
/note="assembly_fragment"	
39396..41939	
/note="assembly_fragment"	
42040..44376	
/note="assembly_fragment"	
44477..47182	
/note="assembly_fragment"	
47283..50373	
/note="assembly_fragment"	
50474..54496	
/note="assembly_fragment"	
54597..57825	
/note="assembly_fragment"	
Query Match	
Best Local Similarity 2.9%; Score 41.6; DB 70; Length 167429;	
Matches 116; Conservative 0; Mismatches 124; Indels 0; Gaps 0;	
QY 1064 attttccatcttccatgcttccatctgatactgtgggtgttgcaattgagacattcat 1123	
DB 158347 ACTTTTCAGCTGCTGATATTTAAAGGTTTTTTTTTTTGTATAGTATAAAATTCAT 158288	
QY 1124 atacaaccagggtgattgttactattgctcttctgtctgtgttggat 1183	
DB 158287 GACCATTTGAAAGGAGAGTGAACACAAATAGACCCATAATCTTCCACGGCAGTACT 158228	
QY 1184 attcatatcacaaaacagagagatctcaacatcacgagagcccttgagaatgaaa 1243	
DB 158227 GCAAAAAAATAAATAAATAAATAAATAAAGAAAAACATAAAATTTCTATTTTAAT 158168	
QY 1244 ctggatcatgataatgctctctgagagatcaagatcacacacacagaacacatagct 1303	
DB 158167 CTTCGATTTGAAATACTGATTATGAAATTTAGACTAAGACACTAATCTAACACTGGT 158108	
RESULT 11	
CNS01RGW 196085 bp DNA HTG 13-MAR-2000	
LOCUS	
DEFINITION Homo sapiens chromosome 14 clone R-355C3, *** SEQUENCING IN PROGRESS ***; 3 ordered pieces.	
AL160231	
ACCESSION AL160231.1 GI:7242513	
VERSION	
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.	
SOURCE human.	
ORGANISM Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 196085)	
REFERENCE	
AUTHORS Direct Submission	
TITLE Submitted (10-MAR-2000) to the EMBL/GenBank/DBJ databases	
JOURNAL IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the	

Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
1 (bases 1 to 22696)
Adams, M. and Venter, J.C.

TITLE	JOURNAL	COMMENT
Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA		This sequence was identified as CD4:10211247 by the submitter. For further information on this sequence you may e-mail to fiv@celera.com .

11yece@red.com.
* NOTE: This is a 'working draft' sequence

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

- * by the finished sequence as soon as it is available and this sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

FEATURES	Location/Qualifiers
1	22605

```
source 1. .22696
```

```

/organism="Prosopbila melanogaster"

```

testabovareim **philosopho** **organi**

```
/db_xref="taxon:7227"
```

Query Match 2.8%; Score 40.8; DB 42; Length 22696;
Best Local Similarity 50.0%; Pred. No. 6.9;
Matches 102; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Oy 1234 agaatggaaacttggatcatgaatatagctccttgaggatcaagaatgacacaccaaga 1293
| | | | | | | | | | | | | | | |
Db 15427 AAAATGTTAGATAAAANAATGAACCAANTCGTAACACTCAATTGTGAATGATCTA 15368

Db 15427 AAAATGTTAGATAAAAAATATGAAACCAATATCGTAACACTCAATTGTTAATGTATCTA 15368

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Qy 1294 aaccatagctagctttgcccaaatctttcatcaacagttttaataattcgtgagggg 1353

Db 15367 ATCCCTATTTCAGTCTCTTTTCCACACTTTCCTTGAACAAATAATGAACTCAAAGCTAGCAGT 15308

DB 15367 ATCCCTATTCAGTCTTTCCACATTTTCGTGGAACAAAAATATGAACCTCAAAAGTAGGAGT 15308

Qy 1354 gagagatcgagataactaattaatggacgtctattatataatgattggaqgttttttqtttta 1413

[illegible]

Search completed: August 1, 2000, 13:36:13
Job time: 57635 sec

Job time: 57635 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: August 1, 2000, 19:50:24 ; Search time 47.63 Seconds
(without alignments)
166.593 Million cell updates/sec
Title: US-09-147-955-10
Perfect score: 1723
Sequence: 1 MSSSSRRRRENEGRMRTL.....GNLDHEYSLLDQDDTPRKP 335
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 188963 seqs, 23686106 residues
Total number of hits satisfying chosen parameters: 188963
Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1723	100.0	335	1 W92951	WO9905287 Seq ID 1
2	166	6.2	402	1 W13437	Plant plastid phos
3	105	6.1	1030	1 W29682	Human thiazide-sen
4	102	5.9	336	1 W27294	Human J59-41 secre
5	98	5.7	375	1 W20731	H. pylori inner me
6	97	5.6	679	1 R10848	Human gibbon ape l
7	97	5.6	679	1 R72969	GALV receptor prot
8	97	5.6	679	1 W70498	Human sodium-lithi
9	97	5.6	679	1 W96988	The gibbon ape leu
10	96.5	5.6	462	1 W97820	Staphylococcus aur
11	95.5	5.5	456	1 P91895	Protein sequence o
12	95	5.5	383	1 W98461	H. pylori GHFO 718
13	94	5.5	457	1 W42430	Escherichia coli g
14	93.5	5.4	312	1 Y11033	H. pylori ORF 04ep
15	93.5	5.4	407	1 Y11034	H. pylori ORF 04ep
16	92	5.3	365	1 R42280	PGE2 receptor codi
17	92	5.3	474	1 W58862	T. halophilus xylu
18	91	5.3	361	1 R42281	PGE2 receptor codi
19	90.5	5.3	415	1 R05539	Rat D2 dopamine re
20	90.5	5.3	415	1 W09388	Mouse dopamine D2
21	90.5	5.3	415	1 Y01598	Rat D2 dopamine re
22	90.5	5.3	444	1 R10544	D2 dopamine recept
23	90.5	5.3	444	1 W09387	Mouse dopamine D2
24	90	5.2	446	1 W98325	Flippase involved
25	88.5	5.1	432	1 W14081	S.thermophilus exo
26	88.5	5.1	467	1 W05741	Presenilin-1-l I14
27	88.5	5.1	473	1 W22180	S.thermophilus exo
28	87.5	5.1	401	1 W92940	Pea Glucose-6-phos
29	87.5	5.1	506	1 W81978	Ehrlichia sp. E74.
30	87.5	5.1	3033	1 R33538	NANBH virus strain
31	86	5.0	365	1 W48086	Human dendritic ce
32	85.5	5.0	348	1 R77224	Hamster ileal/rena
33	85.5	5.0	383	1 W01664	p(rat-edg), G-prot

34	85.5	5.0	383	1 W87791	Rat-edg, G-protein
35	85.5	5.0	681	1 R72970	GLVR-1 protein. Ne
36	85.5	5.0	681	1 W96999	The gibbon ape leu
37	85	4.9	365	1 W8504	Human epidermoid c
38	84.5	4.9	433	1 Y03770	Human G-protein co
39	84.5	4.9	467	1 W05753	Presenilin-1-l E38
40	84	4.9	303	1 W37367	Psbl (Wbpl) protei
41	84	4.9	443	1 R05541	Human pituitary do
42	84	4.9	443	1 R11457	Hman dopamine D2 r
43	84	4.9	443	1 Y01600	Human pituitary do
44	83.5	4.8	467	1 W05743	Presenilin-1-l H16
45	83.5	4.8	482	1 R21409	NADH dehydrogenase

ALIGNMENTS

RESULT 1
ID W92951
AC W92951;
DT 14-MAY-1999 (first entry)
DE WO905287 Seq ID 11.
KW Plant: flavonoid 5-transglycosylation activity; 5TG; variety; colour.
OS Perilla frutescens.
PN WO9905287-Al.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PP 25-JUL-1997; JP-200571.
PR (SUNR) SUNTOY LTD.
PA Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M.
DR WPI; 99-142940/12.
DR N-PSDB; X02830.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
PS Claim 2; Page 81-82; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut
CC flowers harvested from them.
SQ Sequence 335 AA;

Query Match 100.0%; Score 1723; DB 1; Length 335;
Best Local Similarity 100.0%; Pred No. 2.5e-188;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSSSSRRRRENEGRMRTLGLGLGQLYSFDLAIMTFPSASLVSTVDAPLTMSETTVTVV 60
Db 1 MSSSSRRRRENEGRMRTLGLGLGQLYSFDLAIMTFPSASLVSTVDAPLTMSETTVTVV 60
Qy 61 ALLYGTLLYRRHKFLVPWYVALLGFVDVHGNLYLNKAFELTITSVSLDCWTFVWSI 120
Db 61 ALLYGTLLYRRHKFLVPWYVALLGFVDVHGNLYLNKAFELTITSVSLDCWTFVWSI 120
Qy 121 IFTWMFLGTYKSYQVFGAAICVGGLLLVLLSDSGVTAGSNPLLDGFLVITGSLFTLS 180
Db 121 IFTWMFLGTYKSYQVFGAAICVGGLLLVLLSDSGVTAGSNPLLDGFLVITGSLFTLS 180
Qy 181 TVGQECYVKRKDRIEVMVGMVFGMLISATEITVLERNALSSMOWSTGLLAAYVYVALSS 240
Db 181 TVGQECYVKRKDRIEVMVGMVFGMLISATEITVLERNALSSMOWSTGLLAAYVYVALSS 240
Qy 241 FLFCTLTFFLLKKGAAFFNLMLTSDMMAVAIRFTFYNQEDWLYLAFCLVWVGFIY 300
Db 241 FLFCTLTFFLLKKGAAFFNLMLTSDMMAVAIRFTFYNQEDWLYLAFCLVWVGFIY 300
Qy 301 TKTEKDPNTRALENGNLDHEYSLLDQDDTPRKP 335
Db 301 TKTEKDPNTRALENGNLDHEYSLLDQDDTPRKP 335

```

RESULT 2
ID W13497 standard; Protein; 402 AA.
AC W13497;
DE 20-OCT-1997 (first entry)
DE Plant plastid phosphoenolpyruvate-phosphate translocator.
KW Plastid; phosphoenolpyruvate; phosphate; ppt.
OS Brassica oleracea.
PN DE19600357-C1.
PD 13-FEB-1997.
PF 08-JAN-1996; 000357.
PR 08-JAN-1996; DE-000357.
PA (FLUE/) FLUEGGE U.
PA (BADI) BASF AG.
PI Fischer K, Fluegge U, Weber A;
DR WPI; 97-110255/11.
DR N-PSDB; T61549.
PT DNA encoding plant plastid phosphoenolpyruvate-phosphate
PT translocator - for prodn. of transgenic plants, etc.
PS Disclosure: Page 14-16; 17pp; German.
CC The DNA may be used for the prodn. of ppt-expressing pro-or
CC eukaryotic cells; for expression of untranslatable RNA that
CC inhibits the synthesis of endogenous ppt in cells by an
CC antisense effect or by ribozyme activity; for prodn. of
CC DNA sequences encoding a plant plastid ppt with altered
CC specificity; for isolating DNA sequences encoding polypeptides
CC with ppt activity; and for identifying substances that inhibit
CC the transport of phosphoenolpyruvate through inner plastid
CC envelope membranes.
SQ Sequence 402 AA;

Query Match 6.2%; Score 106; DB 1; Length 402;
Best Local Similarity 21.0%; Pred. No. 0.00078;
Matches 68; Conservative 61; Mismatches 133; Indels 16; Gaps 16;

QY 11 ENEGRRRTLLGLGQLVS----FDLAIMTFASLVSTTVADAPLMSFTTYVVAL--- 63
DB 89 EGEKGKTKVLELGLLFAMWYLFNFIYNNQVLEK-ALHAPMTVTLVQFAVGSVLTF 147
QY 64 YGTYLLYRRHFLPWYWYLLGFDVH--GNYLVNKAFELSIYSVILDCWTIWSII 121
DB 148 MWALNLRKPKISAA-QUALPLAVVHTLGNFTNMSLGKVSVSFTTHIKAMEPFSVV 206
QY 122 FTWMLFGTKYSYVQVGAAI-CVGGLLLVLLSDSGVTAAGSNPLGLDFLVIITGLTSL 180
DB 207 LSAMFLGEVPTPW-VIGSIPIVGGVALASVTEVSFNWAG-----FLSAMASNLTNQS 258
QY 181 --TVGQCYVKRKRIEIVAMIGVFGMLISATEITVLERNALSSMQWSTGLLAAYVVAL 238
DB 259 RNVLKRVVWAKDSDLNITLSI-----ITLMSL-----FLMADVTFFS 298
QY 239 SSFLECLTLPFLKMSGAARFNLSMLTSDMWAVAIITFIYNOEVDWL-----YYLA 289
DB 299 EGIKF---TPSVIQSAGV--NVQIVTKSLIALCALCHAY-QQVSMILARVSPVTHSVG 351
QY 290 FC----LVNVGFIYTKTERDPNN 309
DB 352 NCKRVVIVSSVIFFTKTPSPVN 375

RESULT 3
ID W29682 standard; Protein; 1030 AA.
AC W29682;
DE 21-DEC-1998 (first entry)
DE Human thiazide-sensitive NaCl cotransporter TSC.
KW Thiazide-sensitive Na-Cl cotransporter; TSC; human; ion transport;
KW Gitelman's syndrome; Bartter's syndrome; hypokalaemic alkalosis;
KW hypocalcemia; hypomagnesemia; diagnosis; therapy.
OS Homo sapiens.
FH key Location/Qualifiers

136..156
/note="transmembrane domain 1"
166..183
/note="transmembrane domain 2"
219..239
/note="transmembrane domain 3"
257..277
/note="transmembrane domain 4"
286..306
/note="transmembrane domain 5"
340..360
/note="transmembrane domain 6"
376..398
/note="transmembrane domain 7"
453..476
/note="transmembrane domain 8"
508..527
/note="transmembrane domain 9"
531..551
/note="transmembrane domain 10"
561..581
/note="transmembrane domain 11"
585..605
/note="transmembrane domain 12"
W09829431-AL.
09-JUL-1998.
PF 19-DEC-1997; U23553.
PR 31-DEC-1996; US-778052.
PA (UYVA) UNIV VALE.
PI Lifton RP, Simon DB;
DR WPI; 98-388029/33.
DR N-PSDB; V40561.
PT Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl
PT cotransporter genes - useful for developing products for the
PT diagnosis and treatment of ion transport disorders, e.g. Gitelman's
PT Syndrome or Bartter's Syndrome
PS Disclosure: Fig 2; 10pp; English.
CC This is the amino acid sequence of human thiazide-sensitive Na-Cl
CC cotransporter TSC of the renal distal convoluted tubule. Mutations
CC in the TSC gene (see V40561) cause Gitelman's syndrome, a disorder
CC featuring salt wasting and hypokalaemic alkalosis associated with
CC marked hypocalcemia and hypomagnesemia. 17 Different molecular
CC variants inferred to alter the structure of TSC protein were
CC identified in patients with Gitelman's syndrome by SSCP using 27
CC pairs of specific primers (see V40565-618). Identification of the
CC molecular basis of Gitelman's syndrome provides for the genetic
CC diagnosis of this disorder. A method is claimed for determining
CC the presence or absence of a mutated TSC protein conferring altered
CC ion transport by analysing TSC protein samples. Also claimed are
CC vectors, host cells, a method of producing TSC protein, and an
CC antibody that binds altered, but not wild-type, TSC protein. The
CC products and methods can be used for diagnosis of Gitelman's
CC syndrome, to identify carriers and to identify modulators of TSC
CC function of therapeutic value.
SQ Sequence 1030 AA;

Query Match 6.1%; Score 105; DB 1; Length 1030;
Best Local Similarity 21.3%; Pred. No. 0.0039;
Matches 44; Conservative 33; Mismatches 78; Indels 52; Gaps 9;

QY 110 ILDCWTIV-----W-----SIIFTMFLGTYSVYQVGAATCV-----GGLLIV 149
DB 147 MLNINGVILYLRPLWITAQAAGIVLTWIIILLSVTITSITGLSISALSTNGKVKSGTYFL 206
QY 150 LLSDSGVTAAGSNPLGLDFLVIITGLTSTVSG-----QCYCYVKRKRIEIVAMI 200
DB 207 ISRSIGPELGGS---IGLIFAFANAVGVAMHTVGAETVRDLQEQYGAPIDPINDIRII 263
QY 201 GVFGMLISATEITVLERNALSSMQWSTG---LLAAVYVVALSSFLCTLPFLKMSGAA 257
DB 264 GV-----VSVTVLLAISLAGMEMESKAQVFLFLVIMVSFANYLVGTLPSPEDKASKG 316

```

```
QY 258 EFNLMLTSDMNAVAIRTFIYNQEDW 284
DB 317 FFS---YRADI-----FVQNLVFDW 333

RESULT 4
W27294
ID W27294 standard; Protein; 336 AA.
AC W27294;
DT 19-MAY-1998 (first entry)
DE Human J59-41 secreted protein.
KW Secreted protein; anti-inflammatory; immune stimulant; suppressant;
KW human; cytokine; autoimmune disease; regulator; activin; inhibin;
KW J59-41.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..184
FT Protein /label= signal
FT /label= J59-41
FT /note= "secreted protein"
FT
PN W09740151-A2.
PD 30-OCT-1997.
PF 14-APR-1997; U06125.
PR 10-JAN-1997; US-781226.
PR 19-APR-1996; US-635311.
PA (GEMT) GENETICS INST INC.
PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
PI Spaulding V;
DR WPI; 97-535835/49.
DR N-PSDB; T91305.
PT Nucleic acids encoding novel secreted proteins - useful e.g. as an
PT anti-inflammatory, immune stimulant or suppressor, etc.
PS Claim 21; Page 63-64; 81pp; English.
CC This sequence represents a novel secreted protein, J59-41, which is
CC isolated from a clone, J59-41 (ATCC 98028), derived from a human PBMC
CC cDNA library. The nucleic acid and encoded J59-41 secreted protein can
CC be used for research purposes (as markers for tissues, molecular weight
CC markers for gels, primers, probes, etc.), for nutrition (as C, N or
CC carbohydrate source), as a cytokine for cell proliferation and
CC differentiation activity, as immune stimulants or suppressors e.g. for
CC viral, bacterial or fungal infections, for autoimmune diseases such as
CC multiple sclerosis or systemic lupus erythematosus, to regulate
CC haematopoiesis, for tissue growth, as an activin or inhibin, or having
CC chemotactic, chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibitory activities.
SQ Sequence 336 AA;

Query Match 5.9%; Score 102; DB 1; Length 336;
Best Local Similarity 22.3%; Pred. No. 0.0017;
Matches 62; Conservative 41; Mismatches 91; Indels 84; Gaps 13;

QY 52 MSFTYTVVALLGYLLYRRHKFLVPWYVALLGFDVHRGNLYLNKAFELASITSVSIL 111
DB 86 VQAFYSIFGLIELQIQDKRR--IPGKTYMIIAFLVTGTMGLSNTSLGYNYPQVIF 143
QY 112 DCHTWIWSIIFTWFLGKYSYQVYGAACVCG-GLLLVLLSDSGVTAAGSN-----PL 164
DB 144 KCKKLIPVNLGGVFTQGRKYNVAD-VSAACMSGLGLWFTLADS-TTAPNPNRLVLSYS 201
QY 165 LGDFLVTGSLFTSTVCG--EYCVKRRKDRIE-----VAMIGVFGML 206
DB 202 IGPVYILLG-LTCTSGLPATVTCAPNPRTYGYAFLSITGYFGISFVIALIKIFGAL 259
QY 207 ISATEITVLERNALSSMQWSTGLLAAYVVYVYALSFLCTLTTPLLKMGAAFNLSMLTS 266
DB 260 IAVTV-----TTGRKAMTIVL---SFIFFA-KPFTFQ----- 287
QY 267 DMNAVAIRTFIYNQEDWLYLAFCLVVGIF--IYTK 302
DB 288 -----YVWSGLLVVLGIFLNVYSK 306

Query Match 5.7%; Score 98; DB 1; Length 375;
Best Local Similarity 19.9%; Pred. No. 0.0057;
Matches 65; Conservative 57; Mismatches 106; Indels 98; Gaps 18;

QY 52 MSFTYTVVALL--LYGTILLYR-RHKFLVPWYV-YALLGFDVDRHG-----NYLVNK----- 98
DB 72 LSIPTVYFVADMGSVVHWSISMYSGSNDALIQWGHNAVAFVFTSGVIGTIYFLPKESGQ 131
QY 99 ---AFELTSITSVILDCWTFIVWSIIFTWFLGKYSYV---QFVCAATCVGGLLVLL 151
DB 132 PIFSXYKLTIFS-----FWSLMFYIIVAGGHHLIYSTVPDWQVTLSSVFSVVLIL- 180
QY 152 SDSGVTAAGSNPLGLDFLVTGSLFTSTVCGYCYVCKRKDRIEVAMIGVFGML--ISA 209
DB 181 -----PSWGT-----AINMLITMRGOWHQLKESPLIKFLVLAFTFMYLSTLEG 223
QY 210 TETVTLERNALSSM-QWSTGL-----LAAYVVYVYALSFLCTLTTPFLK---MSGAAF-FN 260
DB 224 SIQAIKSVNALAHFTDWIIGHVHDGVLGAVGTFTIASMY-HMTPLFKREIYSGRLVDFQ 282
QY 261 LSMUT-----SDMNAVAIRTFIYNQEDV-----WLY 286
DB 283 FWMTLGLIVLYFSMMIAGITQGMWRDVGNYQFIDTVKALLIPYINIRGVGLMY 342
QY 287 YLAFCLVVGIFITY-----TKTEKDPN 308
DB 343 FIGFIIFAYNIETITAGKKLEREPN 368

RESULT 6
R10848
ID R10848 standard; Protein; 679 AA.
```


CC polypeptides and polynucleotides are useful for diagnosing diseases
 CC related to over or under expression of TR protein by identifying
 CC mutations in the TR gene, or determining TR polypeptide or mRNA
 CC expression levels due to an infection of an organism with the TR
 CC gene (claimed). They can diagnose the stage and type of infection.
 CC TR is also useful for screening for compounds which affect the TR
 CC activity. These can be used in treatment to inhibit antagonist
 CC i.e. antibacterial drugs) or enhance (agonist) TR activity. Direct
 CC administration of TR polypeptides can be used to treat conditions
 CC associated with a lack of TR (claimed). Direct administration of
 CC antisense oligonucleotides prevents expression. TR polypeptides
 CC (administered directly, in a vector (gene therapy) and as a vaccine)
 CC and antibodies induce an immune response to immunise and prevent
 CC disease (claimed). Diseases diagnosed, prevented or treated
 CC include: bacterial, especially *S. aureus*, infections of the
 CC respiratory tract (e.g. otitis media), cardiac (e.g. infective
 CC endocarditis), gastrointestinal (e.g. splenic abscess), CNS (e.g.
 CC cerebral abscess), eye (e.g. conjunctivitis), kidney and urinary
 CC tract (e.g. epididymitis and toxic shock syndrome), skin (e.g.
 CC wound infection), and bone and joint (e.g. septic arthritis). TR
 CC polypeptides, polynucleotides and their (ant)agonists can prevent
 CC adhesion of bacteria to matrix proteins, and are useful for use on
 CC wounds and body implants to prevent bacterial infection. TR is
 CC transcribed during both chronic and acute infections. TR is
 SQ Sequence 462 AA;

Query Match 5.6%; Score 96.5; DB 1; Length 462;
 Best Local Similarity 21.5%; Pred. No. 0.011;
 Matches 56; Conservative 39; Mismatches 121; Indels 45; Gaps 10;
 QY 70 YRRHKELVPWYVALLGFDVHGNYLVNKAFLTSITSVILD--CWTIVMSIIFTWML 127
 DB 17 YSKTKKAVP-----ILFLFVP-SLVINDSKFLISVAIDUNISVTVSQATLAGLVI 70
 QY 128 GTKYSYVQVGAATCV-----GGLLVLLSDSGVTAAGSNPLLDGLFVITGSLFTLST 181
 DB 71 GMAVAVYASLSDAISIRPPPIYGVILLIFGSIIGNIFQSP-----LTLVGRIQTAGL 125
 QY 182 VGQE-----YCVRKRDIEVAMIG-----VFGMLISATIEITVLERNALSSMOWS 226
 DB 126 AAETLVYIVYAKYKSKEDQKTLGLSTSYLSLVIGTLSGGFITYL-----HWT 177
 QY 227 TGLAAVYVVALSFECLTTPFLKMSGAFFNLSMTSDMMAVAITFYINQEVDMY 286
 DB 178 NMFIALIVFTLPFLP-KLLPKENNTNKAHLDFVGLILVATITATVMLFITN--FNWLY 234
 QY 287 YLAFCLVVVGIFIVYTKEDP 307
 DB 235 MIGALIAIVFALYIKNAQRP 255

RESULT 11
 ID P91895
 AC P91895
 DT 17-JUL-1990 (first entry)
 DE Protein sequence of glucuronide permease gene on plasmid pRAJ285
 KW Glucuronide permease gene (GP); glucuronidase (GUS) gene; pRAJ285;
 KW glucuronase; beta-GUS; beta-glucuronides; meladiose permease (MP);
 KW sugar transporters.
 FH Key Location/Qualifiers
 FT region 1..278
 FT /note="coded for by n91708"
 FT region 269..269
 FT /note="coded for by n92949"
 FT misc_difference 279..268
 FT /note="DNA missing from patent"
 PN W08903880-A.
 PD 05-MAY-1989.
 PF 31-OCT-1988; G00936.
 PR 29-OCT-1987; GB-025402.
 PA (JEFF) Jefferson R. A.

PI Jefferson RA;
 DR WPI; 89-150777/20.
 DR N-PSDB; N91708, n92949.
 DR Gene encoding the transport protein glucuronide permease -
 PT allows cellular uptake of beta-glucuronide(s) and detection of
 PT beta-glucuronidase activity.
 PS Disclosure: ; 29pp; English.
 CC It shows homology with lactose permease and mostly with melabiose
 CC permease (MP), which are both sugar transporters. It is very
 CC hydrophobic at regions almost identical to MP. It can be produced by
 CC recombinant techniques or chemically synthesised (including portions or
 CC altered GPs). Also claimed is a recombinant DNA molecule contg. the gene
 CC coding for GP and host cells contg. this molecule. The GP gene is under
 CC the control of a second DNA sequence that regulates expression of GP in
 CC the transformed host. The host cell esp. additionally contains a
 CC glucuronidase (GUS) gene. GP can be used to introduce a substrate for
 CC glucuronase into cells, and GP alters the permeability of live cells.
 CC Expression of GP allows cellular uptake of beta-glucuronides under the
 CC control of virtually any promoter. This system permits the detection of
 CC beta-GUS activity in vivo, and can be conveniently used together with GUS
 CC gene fusions, eliminating the constraints of tissue extracts and
 CC histologic procedures.
 SQ Sequence 456 AA;

Query Match 5.5%; Score 95.5; DB 1; Length 456;
 Best Local Similarity 20.0%; Pred. No. 0.015;
 Matches 76; Conservative 65; Mismatches 136; Indels 103; Gaps 16;
 QY 8 RWRENEGMRRTLLGLGLQGVSDLAIMTFASLSVSTTVADPLTWSETTYTVVALLYGTI 67
 DB 73 RW-----GKFRFLFGTAPLMISVLVFWV---LTDWSHGSKVYVAYLYTMGLGLCISLV 125
 QY 68 LL-----YRRHKPLVPWYVALLGFDVHGNYLVNKAFLTSITS--VSILDC 113
 DB 126 NIPXGSLATAMTQOPQSRARLGAARGIAASLTFVCL--AFLIGPISKNSPEEMVSYVHF 183
 QY 114 WTIVMSI----IFTMFLGTYKSYQVFGA-----AICVGGLLLVLS 152
 DB 184 WTIVLAGNVLYIFCFKSTRENVVRIVQAQPSLNISLQTLKRNPLFMCIGA-LCVLIS 242
 QY 153 DSGVTAAGSNPLLDGLFVITGSLFTLSTVGQE---YC-----VRRKRDIEV 197
 DB 243 TFAVSASS---LFYVRYVINDTGLFTVLVQNPQWYCGITGAGAGWVARIGKNTFLIG 299
 QY 198 AMIGVFGML-----ISATEITVLERNALSSMOWSTGLAA 232
 DB 300 ALLGTCGYLLFFWVSWSLPVALVALAIASIGGVMTVMWALEADTVEYGEYLTGVRIE 359
 QY 233 YVYVALSSF-----LFCITLTPFLKMSGAFFNLSMTSDMMAVAITFYINQEVDM 285
 DB 360 GLTYSLSFSFTRKCGAIGGSIAPILGLSG---YIANQVQTPVEVINGIRTSLALVPCGFM 416
 QY 286 YLAFCLVVVGIFIVYTKTEK 305
 DB 417 -LLAPVJ----INFVPLTDK 431

RESULT 12
 ID W98461
 AC W98461;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHP0 718 protein.
 KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 OS Helicobacter pylori.
 PN W09843478-A1.
 PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.

CC also of use in screening for compounds having the ability to interfere
CC with the *H. pylori* life cycle or to inhibit *H. pylori* infection.
SQ Sequence 312 AA;

CC when the n: pY101
SQ Sequence 312 AA;

```

Query Match      5.48; Score 93.5; DB 1; Length 312;
Best Local Similarity 20.2%; Pred. No. 0.014;
Matches 54; Conservative 46; Mismatches 112; Indels 55; Gaps 10;

QY 18 TLIGLGLGOLVDFDLAIMTFSSASIVSTVDA--PLTMSFTTYTVVALLYTYTILYRR--- 72
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 TTIGPIFGSLIFSATKTSNLSLIDKLADAKSVQMPGLGAVFSELLAVMYLLKPDV 111
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 -----HKELVPWYKALLGDFVDHGNLVNKNKAPELTSTSVSILDCWIV 117
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 EKEMPKETTKQLSFHKHFV-----FGALGIFVYVGGVGAIGSELVLVSFEKLLNLDAQSSA 167
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 WSIIFTW-----MFLG-----KYSYIQVGAACVGGGLLVLLTSDSGVTAAGSN 162
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 HVLVYVWYGAMVREGLSGALMKNKIAPNKLAFNALSSIIILIALAILI-----GKIALPAL 223
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 PLLGDFLVITGSIPLTFLTVGQECVVKORIEVAMIGVFGMLISATE-----ITVLE 216
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 TVGPFNSIMFTFISLATLNLGHTSKASGVISMAIG--GALIPPIQGVVTDMLTATE 281
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 RNALSSMOWSTGLLAAYV--YALSSF 241
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 SNLLXA--YSPVLLCYFYILFFALKGY 306
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

Y11034	
ID	Y11034 standard; Protein; 407 AA.
AC	Y11034;
DE	08-JUN-1999 (first entry)
DT	H. pylori ORF 04ep1903-19699182-cl_43 inner membrane protein.
FE	Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW	secreted protein; cytoplasmic protein; cellular protein.
KN	Helicobacter pylori.
OS	W09824475-A1.
PN	11-JUN-1998.
PD	05-DEC-1997; U22104
PF	14-JUL-1997; US-891928.
PR	05-DEC-1996; US-759625.
PR	25-MAR-1997; US-823745.
PT	(ASTR) ASTRA AB.
PI	Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
PI	WPI: 98-333051/29.
DR	N-PSDB: X30563.
PT	New isolated Helicobacter pylori nucleic acids - used to develop
PT	products for the diagnosis, prevention and treatment of infection by
PT	H. pylori and other Helicobacter species
PS	Claims 37, 41; Page 198-199; 33pp: English.
CC	Recombinant or substantially pure preparations of H. pylori polypeptides
CC	are disclosed, together with the nucleic acids encoding them. In all,
CC	97 ORFs are shown. The proteins are variously cell envelope proteins,
CC	cytoplasmic proteins, secreted proteins or other cellular proteins.
CC	Vaccines containing the nucleic acids or proteins are claimed, as are
CC	probes containing at least 8 nucleotides from the nucleic acid
CC	sequences. The vaccines are useful for treating or reducing the risk of
CC	H. pylori infections, and the probes can be used diagnostically for
CC	detecting the presence of Helicobacter in a sample. The products are
CC	also of use in screening for compounds having the ability to interfere
CC	with the H. pylori life cycle or to inhibit H. pylori infection.
CC	Sequence 407 AA;
CC	

Query Match 5.4%; Score 93.5; DB 1; Length 407;
Best Local Similarity 20.2%; Pred. No. 0.021;
Matches 54; Conservative 46; Mismatches 112; Indels 55; Gaps 10;

Db	147	TTLGPIFGSLLIFSATKTSNDLSLIDKLADAKSVQMPYLG LAVFSLLALWYLLKLPDV	206
QY	73	-----HKFLPWPWYALLGFDVHGNYLVNKAFELTSITSYSLDCWTIV	117
Db	207	EXEMPKETQKSLFSHKHFV-----FGALGIFVYGVGEVAIGSFVLSPEKLLNLDQA	262
QY	118	WSLIITW-----NPLGT-----KYSVYQFVGAAICVGGLLVLVLSDSGVTAAGSN	152
Db	263	HVLVYVWGGMVGRFLGSLAMNKNIAPNKYLFAPNALSSIIIALATLI-----GGXIALFAL	318
QY	163	PLLGDPFLVITGSILFTSLTSGOEYCVYKRRKDRIEWAAMIGVFGMLISATE-----ITVLE	216
Db	319	TFVGFFNSINFTFISPLATLMLGHLITSKASGVISMNIG--GALIPPIQGVVTDMLTATE	376
QY	217	RNALSGMOWSTGLLAAYVV--YALSSF	241
Db	377	SNLLYA--YSVPLLCYFYILFFALKGY	401

Search completed: August 1, 2000, 19:50:26
Job time: 21506 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 19:51:34 ; Search time 59.57 Seconds
(without alignments)
86.206 Million cell updates/sec

Title: US-09-147-955-10
Perfect score: 1723
Sequence: 1 MSSSSRRWRENEGMRRTLL.....GNLDHEYSILLEDDDTPRKP 335

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues
Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5a_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5a_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/5a_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/backfiles!.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103.5	6.0	426	5	Patent No. 5268463-8
2	97	5.6	680	1	US-07-674-287B-2
3	97	5.6	680	2	US-08-436-900A-2
4	96.5	5.6	462	2	US-08-898-976-2
5	96.5	5.6	462	2	US-08-898-976-4
6	95.5	5.5	456	5	Patent No. 5432081-7
7	94	5.5	457	2	US-08-882-704A-6
8	92	5.3	365	1	US-08-390-162-2
9	92	5.3	365	1	US-08-685-945B-2
10	91.5	5.3	415	1	US-08-194-338-10
11	91	5.3	361	1	US-08-390-162-4
12	91	5.3	361	1	US-08-685-945B-4
13	90.5	5.3	214	3	US-08-807-300-3
14	90.5	5.3	238	2	US-08-103-170-13
15	90.5	5.3	303	1	US-08-196-989B-9
16	90.5	5.3	303	2	US-08-760-936-9
17	90.5	5.3	444	1	US-07-781-254A-1
18	90.5	5.3	462	3	US-08-788-231A-15
19	88.5	5.1	444	1	US-07-626-618A-19
20	88.5	5.1	444	1	US-08-333-977-19
21	88.5	5.1	473	1	US-08-597-236-13
22	88.5	5.1	473	1	US-08-746-682A-13
23	87.5	5.1	3033	1	US-07-925-695-5
24	86.5	5.0	443	1	US-07-626-618A-18
25	86.5	5.0	443	1	US-08-333-977-18
26	86	5.0	385	1	US-08-416-756A-3
27	85.5	5.0	348	1	US-08-176-126B-2
28	85.5	5.0	348	2	US-08-669-435-2

29	85.5	5.0	348	4	PCT-US94-14431A-2	Sequence 2, Appli
30	85.5	5.0	371	2	US-08-928-692-20	Sequence 20, Appli
31	85.5	5.0	383	1	US-08-196-989B-4	Sequence 4, Appli
32	85.5	5.0	383	2	US-08-760-936-4	Sequence 4, Appli
33	85.5	5.0	682	2	US-08-436-900A-4	Sequence 4, Appli
34	85	4.9	341	2	US-08-846-762-92	Sequence 92, Appli
35	84.5	4.9	433	2	US-08-919-624-1	Sequence 1, Appli
36	84.5	4.9	438	2	US-08-677-049-9	Sequence 9, Appli
37	84	4.9	303	2	US-08-846-762-20	Sequence 20, Appli
38	84	4.9	303	2	US-08-846-762-91	Sequence 91, Appli
39	84	4.9	457	5	5268463-7	Patent No. 5268463
40	83	4.8	184	2	US-08-602-359A-33	Sequence 33, Appli
41	82.5	4.8	388	3	US-08-155-005A-6	Sequence 6, Appli
42	82.5	4.8	407	2	US-08-875-972-4	Sequence 4, Appli
43	82.5	4.8	428	5	5432081-9	Patent No. 5432081
44	82.5	4.8	434	2	US-08-529-600B-2	Sequence 2, Appli
45	82.5	4.8	434	2	US-08-973-275-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
5268463-8
; Patent No. 5268463
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: PLAANT PROMOTER a-GLUCURONIDASE GENE
; CONSTRUCT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,976
; FILING DATE: 08-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; SEQ ID NO:8:
; LENGTH: 426
5268463-8

Query Match 6.0%; Score 103.5; DB 5; Length 426;
Best Local Similarity 21.1%; Pred. No. 0.018;
Matches 80; Conservative 63; Mismatches 133; Indels 103; Gaps 18;

QY	8	RWRENGEMRTLLGLGLGQLYSFDLAIMTESASLVSTTVDAPLTMSFTTVVVALVGTI	67
DB	65	RW-----GKFRPFLFGTAPLMIFD-SVLVFWY-LTDSHGSKVYVYALTGTAGLCYSLV	118
QY	68	LL-----YRRHKFLVPWYVALGFDVHGNLYLVNKAFLTSLTS-VSLDC	113
DB	119	NIPYGLSATAMTQQPSRARLGAARGIAASLTFVCL--AFLIGPSIKNSPEEMVSYHF	176
QY	114	WTIVWSI---LFTWNEFLGTYVYQFVGA-----AICVGGLLLVLS	152
DB	177	WTIVLAIGMWLYFCFKSTRENVRIVQAQPSLAISLQTLKRNPLFMLIGA-LCVLIS	235
QY	153	DSGVTAAAGNPLGLDFLVITGSLFTLTSTVQGE---YC-----VKR--KDRIEVVA	198
DB	236	TFAYASSSL-----FYVLNDTGLFTVLVLVQNPFCGIGTAGAXMVARIGKTKFLIGA	289
QY	199	MIGVGMCL-----ISATEITVLENNLSSHOWSTGTLAAV	233
DB	290	LLGTGCLLFFWVWSLPAVALAIASIGQVMTVMVALEADTVGEYLTGVRIEG	349
QY	234	VVALSSP-----LFTCLTFFLLKNSGAFFNLMLTSDMVAIVRTFIYVQEDWLY	286
DB	350	LYSLFSFTRKCGAIGGSIPAFILGLSG---YIANOVQTPVEIMGIRTSIALVPGCFM-	405
QY	287	YLAFCLVVVGIFVYTKTEK	305
DB	406	LLAFVII---WEFIPLTDK	421

RESULT 2
US-07-674-287B-2
; Sequence 2, Application US/07674287B
; Patent No. 5414076
; GENERAL INFORMATION:
; APPLICANT: Bryan Mark O'Hara
; TITLE OF INVENTION: Gibbon Ape Leukemia
; TITLE OF INVENTION: Virus Receptor
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Karen A. Lowney
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: CT
; COUNTRY: USA
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII converted from IBM DW4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/674,287B
; FILING DATE: 19910325
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A., Dr.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 31,104-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2361
; TELEFAX: 203 321 2971
; TELEX: 710 474 4059
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 680 Amino Acid Residues
; TYPE: AMINO ACID SEQUENCE
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Protein
US-07-674-287B-2

Query Match 5.6%; Score 97; DB 1; Length 680;
Best Local Similarity 21.7%; Pred. No. 0.15;
Matches 66; Conservative 49; Mismatches 105; Indels 84; Gaps 16;
QY 76 LVPWYVALLGFDVHGNVLYNKAPELT-SITSVSILDCW-TIVWSIITFMFLGTRYSV 133
DB 18 LVDYLMMLILGFI-----IAFVLAFSGANDVANSFGTAVGSGVVTLKQACILASI 68
QY 134 YQVGAALCVG-----CLLLVLL--SDSGVTAAGS-----NPLL 165
DB 69 PETVGSVLLGAKVSETIRKGLIDVEMNSTOGLLMAGSVAMFGSAVHQLVASFLKLPIS 128
QY 166 GDFLVITGSILFTLSTVGQCYVKRKDRIEV-----AMIGVF-GMLISATETVLER 217
DB 129 GTHCIVGATIGFSLVAKGQE-GVKWSELKIVMSFVSPLLSGIMSGILFPLVRAFIHK 187
QY 218 NALSSMQWSTGLLAAYVYVALSSFLCT--LTPFLKMSGAAFFNLSLMTSDMNAVAIRT 275
DB 188 -----ADVPNGRLALPVFA-----CTVGINLFSIMYTGAPLLGFDKL--PLMGTILIS 235
QY 276 FIYNQVDWLYLAFLCVLVVVGIFVYTKTEK-----DPNTRALENGNLDHEYSLLED 327
DB 236 -----VGCVAFFCALIWFVFCPRMKRKIEREIKCSPSPLEMKKN-----SLKED 281

QY 328 QDDT 331
DB 282 HBET 285
RESULT 3
US-08-436-900A-2
; Sequence 2, Application US/08436900A
; Patent No. 5874264
; GENERAL INFORMATION:
; APPLICANT: O'Hara, Bryan M.
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,900A
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.
; REGISTRATION NUMBER: 31,088
; REFERENCE/DOCKET NUMBER: 31,104-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-683-2158
; TELEFAX: 201-683-4117
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 680 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-436-900A-2

Query Match 5.6%; Score 97; DB 2; Length 680;
Best Local Similarity 21.7%; Pred. No. 0.15;
Matches 66; Conservative 49; Mismatches 105; Indels 84; Gaps 16;
QY 76 LVPWYVALLGFDVHGNVLYNKAPELT-SITSVSILDCW-TIVWSIITFMFLGTRYSV 133
DB 18 LVDYLMMLILGFI-----IAFVLAFSGANDVANSFGTAVGSGVVTLKQACILASI 68
QY 134 YQVGAALCVG-----GLLLVLL--SDSGVTAAGS-----NPLL 165
DB 69 PETVGSVLLGAKVSETIRKGLIDVEMNSTOGLLMAGSVAMFGSAVHQLVASFLKLPIS 128
QY 166 GDFLVITGSILFTLSTVGQCYVKRKDRIEV-----AMIGVF-GMLISATETVLER 217
DB 129 GTHCIVGATIGFSLVAKGQE-GVKWSELKIVMSFVSPLLSGIMSGILFPLVRAFIHK 187
QY 218 NALSSMQWSTGLLAAYVYVALSSFLCT--LTPFLKMSGAAFFNLSLMTSDMNAVAIRT 275
DB 188 -----ADVPNGRLALPVFA-----CTVGINLFSIMYTGAPLLGFDKL--PLMGTILIS 235
QY 276 FIYNQVDWLYLAFLCVLVVVGIFVYTKTEK-----DPNTRALENGNLDHEYSLLED 327
DB 236 -----VGCVAFFCALIWFVFCPRMKRKIEREIKCSPSPLEMKKN-----SLKED 281
QY 328 QDDT 331

[illegible][illegible]

RESULT 6

5432081-7
; Patent No. 5432081
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI
; GLUCONIDE PERMEASE GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,546
; FILING DATE: 15-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 447,976
; FILING DATE: 08-DEC-1989
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; SEQ ID NO: 7:
; LENGTH: 456
5432081-7

Query Match 5.5%; Score 95.5; DB 5; Length 456;
Best Local Similarity 20.0%; Pred. No. 0.12;
Matches 76; Conservative 65; Mismatches 136; Indels 103; Gaps 16;
QY 8 RWRENEGMRRTLLGLGLGQLVSPDLAIMTFSASLVSTTVDPAPLTMSTFTYTVVALLXGTI 67
DB 73 RW-----GKRPFLFGTAPLMIFSVLVFWV---LTDWSHGSKVYVAYLTYMGLGLCYSLV 125
QY 68 LL-----YRRHKFLVPWYVYALLGFVDVHGNYLVNKAFLTSITS-VSILDC 113
DB 126 NIPYGLSATANTQOQSRARLGAARGIASLTFVCL--AFLGPIKSNSSPEEMSVYHF 183
QY 114 WTIVWSI-----IFTWMLGTYKYVYQFVGA-----AICVGGLLLVLS 152
DB 184 WTIVLAGMVLTYPCFKSTRENVIRIVAQPSLNISLOTLRNRPLFLMCLIGA-LCVLIS 242
QY 153 DSGVTAAGSNPLGLDFLVINGSLITSTVGQE---YC-----VKKRDRIEV 197
DB 243 TFAVSASS---LFVVRVLDNTGLFTVLVQVQWYCGIGTAGAXMVARICKNTFLIG 299
QY 198 AMIGVFCML-----ISATEITVLERALSSQWSTGLLAA 232
DB 300 ALLTCGYLLFFFWVNSLPLVALALAIASIGQGVMTVMVMALEADTVVEYGEYLTGVRIE 359
QY 233 VYVVALSSF-----LFCITLTPFLKMSGAAFFNLMLTSDMVAIVARTIYINQVDWL 285
DB 360 GLTYSLSFTRKCGQAGSIPAFILGLSG---YIANQVQTPVEVINGIRTSIALVPCGFM 416
QY 286 YLAFCLVVGVIYFTYKTEK 305
DB 417 -LLAFVI-----INFYPLTDK 431

RESULT 7

US-08-882-704A-6
; Sequence 6, Application US/08882704A
; Patent No. 5879906
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Wilson, Katherine J.
; APPLICANT: Leader, Michael
; TITLE OF INVENTION: GLUCONIDE REPRESSORS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,704A
; FILING DATE: 25-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5879906tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 190106.404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-704A-6

Query Match 5.5%; Score 94; DB 2; Length 457;

Best Local Similarity 18.8%; Pred. No. 0.17;
Matches 82; Conservative 71; Mismatches 125; Indels 158; Gaps 18;
QY 17 RFLGLGLGQ-----LVSPDLAIMTFSASLVSTTVDPAPLTMSTFTYTVV--- 60
DB 8 RTIVGSLGDVANNFAFMGALFLLSYTDOVAGVAAAGTML--LLVRVDFADVFAG 65
QY 61 -----ALLYGT--LLYRRHKFLVPWY-----YALLGFVDVHGNT-L 95
DB 66 RVYDSVNTRWKFRPFLFGTAPLMIFSVLVFWPTDWSHGSKVYVAYLTYMGLGLCYSL 125
QY 96 VNAKEP--LTSITS-----VSILDCW 114
DB 126 VNIPYGLSATANTQOQSRARLGAARGIASLTFVCLAFGLGPIKSNSSPEEMSVYHF 185
QY 115 TIWWSI-----IFTWMLGTYKYVYQFVGA-----AICVGGLLLVLS 153
DB 186 TIVLAGMVLTYPCFKSTRENVIRIVAQPSLNISLOTLRNRPLFLMCLIGA-LCVLIST 244
QY 154 SCVTAAG-----SNPLGLDFLVITGSLITSTV---GQEYCVKRDRIEVAMIG 201
DB 245 FAVSASSLFYVRYVLDNTGLFTVLVQVQWYCVASAPLVPGMVARICKNTFLIGALLG 304
QY 202 VFGML-----ISATEITVLERALSSQWSTGLLAAVYVY 236
DB 305 TCGYLLFFFWVNSLPLVALALAIASIGQGVMTVMVMALEADTVVEYGEYLTGVRIEGLTY 364
QY 237 ALSSP-----LFCITLTPFLKMSGAAFFNLMLTSDMVAIVARTIYINQVDWL 289
DB 365 SLFSETRKCGQAGSIPAFILGLSG---YIANQVQTPVEVINGIRTSIALVPCGFM-L 420
QY 290 FCLVVGVIYFTYKTEK 305
DB 421 FVI-----INFYPLTDK 432

RESULT 8
US-08-390-162-2
; Sequence 2, Application US/08390162
; Patent No. 5576192
; GENERAL INFORMATION:
; APPLICANT: Ichikawa, Atsushi
; APPLICANT: Narumiya, Shuh
; TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390.162
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
APPLICATION NUMBER: JP 036580-1992
FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-390-162-2

Query Match 5.3%; Score 92; DB 1; Length 365;
Best Local Similarity 19.7%; Pred. No. 0.2;
Matches 68; Conservative 49; Mismatches 114; Indels 114; Gaps 14;
QY 4 SSSRRRNEGMRRTLLGLG-----LGQVSDLAIMTFASLSVSTTTVDAP-----49
DB 51 SRSYRRRESKRKSFLLCIGWLALDVLGQLTSPVILVLSQRRWEQDPSGRLCFTFF 110
QY 50 -LWMSSTTVVVALYGLILYRRHKFLVPWYTA-----LLGFVDVHGNYLVNK 98
DB 111 GLTM--TVFGLSSLVASAMVERALAIAPHWYASHMKTRATPVLLGV-----WLSVL 162
QY 99 AFELTSITSVILDCWTIYWSIIFTMFLGT-----KYSYTFQVGAACVGGLL 147
DB 163 AFALLPVLGVG--RYSVQWP--GTWCFISTGPAGNETDPAREPGSVAFASAFACGLLA 217
QY 148 LVLLSDSGVTAAGSNPLLGDFLVTGSLTFLSTVQGEYCVKR-KDRIEVVAMIGVGM 206
DB 218 LV-----VTFACNLATIKALVSRCAKAAVSQSSAQMGRI 252
QY 207 ISATEITVTLERNALSSMOWSTGLLAAYVYVYALSSFLCTLTFFLLKMSGAFFNLN---263
DB 253 TTETAQLMGIMCVLSVCS-----PULLIMLMKMFNQMSVEQC 291
QY 264 -----LISDMWAVAIRFTIYNQEVQD-WLYYL-----AFCLV 293
DB 292 KTMGKEKCNFSLIAVRLASUNQLDPPWYLLLRKILLRKPCQI 336

RESULT 9
US-08-685-945B-2
; Sequence 2, Application US/08685945B
; Patent No. 580415

GENERAL INFORMATION:
APPLICANT: Ichikawa, Atsushi
APPLICANT: Narumiya, Shuh
TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685.945B
FILING DATE: 22-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 036580-1992
FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-685-945B-2

Query Match 5.3%; Score 92; DB 1; Length 365;
Best Local Similarity 19.7%; Pred. No. 0.2;
Matches 68; Conservative 49; Mismatches 114; Indels 114; Gaps 14;
QY 4 SSSRRRNEGMRRTLLGLG-----LGQVSDLAIMTFASLSVSTTTVDAP-----49
DB 51 SRSYRRRESKRKSFLLCIGWLALDVLGQLTSPVILVLSQRRWEQDPSGRLCFTFF 110
QY 50 -LWMSSTTVVVALYGLILYRRHKFLVPWYTA-----LLGFVDVHGNYLVNK 98
DB 111 GLTM--TVFGLSSLVASAMVERALAIAPHWYASHMKTRATPVLLGV-----WLSVL 162
QY 99 AFELTSITSVILDCWTIYWSIIFTMFLGT-----KYSYTFQVGAACVGGLL 147
DB 163 AFALLPVLGVG--RYSVQWP--GTWCFISTGPAGNETDPAREPGSVAFASAFACGLLA 217
QY 148 LVLLSDSGVTAAGSNPLLGDFLVTGSLTFLSTVQGEYCVKR-KDRIEVVAMIGVGM 206
DB 218 LV-----VTFACNLATIKALVSRCAKAAVSQSSAQMGRI 252
QY 207 ISATEITVTLERNALSSMOWSTGLLAAYVYVYALSSFLCTLTFFLLKMSGAFFNLN---263
DB 253 TTETAQLMGIMCVLSVCS-----PULLIMLMKMFNQMSVEQC 291
QY 264 -----LISDMWAVAIRFTIYNQEVQD-WLYYL-----AFCLV 293

Db 292 KTMGKEKCNFLLAVLASLQILDPWVYLLLRKILLRKFCQI 336

RESULT 10
US-08-194-338-10
; Sequence 10, Application US/08194338
; Patent No. 5474898
; GENERAL INFORMATION:
; APPLICANT: Venter, John C.
; APPLICANT: Fraser, Claire M.
; APPLICANT: McCombie, William R.
; TITLE OF INVENTION: OCTOPAMINE RECEPTOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/194,338
; FILING DATE: 08-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/676,174
; FILING DATE: 28-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH101.001DVI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
US-08-194-338-10

Query Match 5.3%; Score 91.5; DB 1; Length 415;
Best Local Similarity 19.5%; Pred. No. 0.26;
Matches 64; Conservative 36; Mismatches 95; Indels 133; Gaps 14;

QY 78 PWYV-----ALIGFVDVHGNYLV-----NKAFLTS---ITSVSILDC-----WT 115
DB 32 PHNYAYMLTLFLIFVGNVLYCHVAVSEKALQTTWYLVLSAVADLLVAVLVWVW 91
QY 116 IVWSIIFTWFLGTYKSVYQFVGAAICVGGLLLVLSDSGVTAAGSNPLLGDFLVITGSI 175
DB 92 VYLEVVGSEKFSRIHCDF-----VTL-----DVNMCTASI 122
QY 176 LFTLSTVGQEVCKVRKDRIEVAMIGVGMGISATEITVLERNALSSMOWSTGLLAAYV 235
DB 123 L-----NLCATSIDRYTAVM-----PMLYTRSSRRRTVMIA--IV 159
QY 236 YALSSFLCTTTPFLKMSGAFFNLSMTSDMVAIVARTFYNOEVDWLY--YLAFLV 293
DB 160 WYLSFTISCPL-----IFQLNNTDQECIIANPAFYVSSIVSFVFPFVITLV 208
QY 294 VVGIFIVTKTEKOPNNTR-----ALENGNLD----- 319

Db 209 YKIYIVLRKKRVNKRSSRAFRANLKTPLKDAARAGELMEMLSSTAPPERTYSP 268
QY 320 -----HEVSL-----LEDQDDTPRKP 335
DB 269 IPPSHHQTLTLPDPSSHGLHSNPDSAPK 296

RESULT 11
US-08-390-162-4
; Sequence 4, Application US/08390162
; Patent No. 5576192
; GENERAL INFORMATION:
; APPLICANT: Ichikawa, Atsushi
; APPLICANT: Narumiya, Shuh
; TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,162
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024179
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: JP 036580-1992
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 064889-1992
; FILING DATE: 23-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 04221-0020-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-390-162-4

Query Match 5.3%; Score 91; DB 1; Length 361;
Best Local Similarity 19.8%; Pred. No. 0.25;
Matches 68; Conservative 48; Mismatches 113; Indels 114; Gaps 14;

QY 4 SSSRRWRNECMRRITLLGLG-----LGQVSEFDLAIMTFSASLVSTTVDPAP----- 49
DB 51 SRSYRRSRKRRKFKFLICIGLWLTDLVGQLTSPVVLVLSQRWQLDPSGLCTFF 110
QY 50 -LTSFTTYTVVALLXGTYLLYRRHKFLVPWWYA-----LLGFVDVHGNYLVNK 98
DB 111 GLTM--TVFGLSLLVASAMAVERALAIRAPHWYASHMKTRATPVLLGV-----WLSVL 162
QY 99 AFELTSITSVSILDCWITVWSIITFMFLGT-----KYSYQFVGAAICVGLL 147

Db 163 AFALLPVLGVG---RYSQWP---GTWCFTSTGPNAGNETDPAREPGSVAFASAFACLGLLA 217
QY 148 LVLLSDSGVTAAGSNPLLDGFLVITGSLFTLSTVGQYCVKR-KDRIEYVAMIGVFGML 206
Db 218 LV-----VTFACNLATIKALVSRCAKAAVSQSSAQMGR 252
QY 207 ISATEITVLERNALSSMOWSTGLLAAVYVYALSFLCTLTPFLKMSGAAPNLSM--- 263
Db 253 TTETAQLMGIMCVLVCWS-----PLLIMLMKMFNQMSVEQC 291
QY 264 -----LTSMDMVAIRTFYINQEVW-WLYYL-----AFC 291
Db 292 KTMQKEKCNFLIAVRLASLNQILDPEWVYLLLRKILLRKC 334

RESULT 12
US-08-685-945B-4
; Sequence 4, Application US/08685945B
; Patent No. 5804415
; GENERAL INFORMATION:
; APPLICANT: Ichikawa, Atsushi
; APPLICANT: Narumiya, Shuh
; TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/685,945B
; APPLICATION DATE: 22-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024179
; FILING DATE: 23-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 036580-1992
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 064889-1992
; FILING DATE: 23-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 04221-0020-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-685-945B-4

Query Match 5.3%; Score 91; DB 1; Length 361;
Best Local Similarity 19.8%; Pred. No. 0.25;
Matches 68; Conservative 48; Mismatches 113; Indels 114; Gaps 14;

QY 4 SSSRRWRNEGRRRTLLIGG-----LGQLVSPDLAIMTFASLSTVTDAP----- 49
Db 51 SRSYRRRSRRKRSFLCIGLWALDLVQGLTSPVILVILSQRWQDPSGRLCRTFF 110

QY 50 -LTMSTTTVVVALLYGILLYRRHKFLVPWYTA-----LLGFVDVHGNTLVNK 98
Db 111 GLTM--TVFGLSLLVASAMAVERALAIRAPHWYASHMKTRATPVLGV-----WLSVL 162
QY 99 AFLTITSVSILDCWTVIWSLIITFMELGT-----KYSYQFYGAACVGGLL 147
Db 163 AFALLPVLGVG---RYSQWP---GTWCFTSTGPNAGNETDPAREPGSVAFASAFACLGLLA 217
QY 148 LVLLSDSGVTAAGSNPLLDGFLVITGSLFTLSTVGQYCVKR-KDRIEYVAMIGVFGML 206
Db 218 LV-----VTFACNLATIKALVSRCAKAAVSQSSAQMGR 252
QY 207 ISATEITVLERNALSSMOWSTGLLAAVYVYALSFLCTLTPFLKMSGAAPNLSM--- 263
Db 253 TTETAQLMGIMCVLVCWS-----PLLIMLMKMFNQMSVEQC 291
QY 264 -----LTSMDMVAIRTFYINQEVW-WLYYL-----AFC 291
Db 292 KTMQKEKCNFLIAVRLASLNQILDPEWVYLLLRKILLRKC 334

RESULT 13
US-08-807-300-3
; Sequence 3, Application US/08807300
; Patent No. 6010852
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PROTON ATPASE SUBUNIT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,300
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0222 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1132516
; US-08-807-300-3

Query Match 5.3%; Score 90.5; DB 3; Length 214;
Best Local Similarity 21.5%; Pred. No. 0.14; 59; Indels 81; Gaps 11;
Matches 47; Conservative 32; Mismatches 32

QY 51 TMSFTTYYVALLGTYLLY-----RRHKFLVPWY-----WYAL-LGF----- 87
DB 10 TATLTVTITLIL-GTGLFYMLSGQGHREFDIGNFLTSTSPHWMAGIGIGFSLSLVLGAG 68
QY 88 --VDVHGNYLVN---KAFELTSITSVILDCWTI-VMSIIFTWMLFG----- 128
DB 69 WGIPTGSSILGGVKAPRIRTKNLVSIIFCEAVAIFGIIMAFVFGVGLAEFRREDLPDT 128
QY 129 -----TKSYQYQVGAICVG-----GLLLVLLSDSGVTAAGSNPLLDGLVITG 173
DB 129 EDGMAILARNLASGYMIFGGGLVGLVNLVCGLVAGVIGSGAAIADAANPALFVKILI-- 186
QY 174 SILFTLSTVGQYCVKRRKRIEYVA-MIGVFGMLISATE 211
DB 187 -----IEIPASAIGLFGMLIGIVQ 205

RESULT 14
US-08-103-170-13
; Sequence 13, Application US/08103170
; Patent No. 5885824
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Recombinant Genomic Clones Encoding
; TITLE OF INVENTION: Histamine H1, H2, and H3 Receptors, Methods For Production
; TITLE OF INVENTION: Thereof, and Proteins Encoded Therefrom
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCELLEND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/103,170
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/633,060
; FILING DATE: 24-DEC-1990
; NAME: Lavalleye, Jean-Paul
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-017-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 24885 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: rat
US-08-103-170-13

Query Match 5.3%; Score 90.5; DB 2; Length 238;
Best Local Similarity 20.9%; Pred. No. 0.16;
Matches 54; Conservative 32; Mismatches 85; Indels 87; Gaps 11;
QY 78 PWYV-----ALLGFDVHGNYLV-----NKAFELTS---ITSVSILDC-----WT 115

DB 32 PHNYTYAMLLTLIIIFVGNVLVCMVAVSREKALQTTTNYLVSLAVADLLVATLVMPV 91
QY 116 IWMSIIFTWMLFGTKSYQYQVFGAAICVGGLLLVLLSDSGVTAAGSNPLLDGLVITGSI 175
DB 92 VYLEVVGEMKFSRIHCDIF-----VTL-----DVMCTASI 122
QY 176 LETLSTVGOEYCVKRRKRIEYVAMIGVFGMLISATITVLERNALSSMOWSTGLLAAYV 235
DB 123 L-----NICATSIDRYTAVAM-----PMLNTRYSSKRRTVMIA--IV 159
QY 236 YALSSFLFCTLTPELLKMSGAAPFNLSMLTSDMWAVAIRTFYINQVVDWLY--YLAFLCV 293
DB 160 WYLSFTISCLP-----LFLNNTQDNECIIANPAFVYVSIIVSIFVYVFTVTLV 208
QY 294 VVGIEIYTKTEKDPNTR 311
DB 209 YIKIYIVURKKRVNTR 226

RESULT 15
US-08-196-989B-9
; Sequence 9, Application US/08196989B
; Patent No. 5585476
; GENERAL INFORMATION:
; APPLICANT: MacLennan, A. John
; TITLE OF INVENTION: Molecular Cloning and Expression of
; TITLE OF INVENTION: G-Protein Coupled Receptors
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/196,989B
; FILING DATE: 15-FEB-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: MAC-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-196-989B-9

Query Match 5.3%; Score 90.5; DB 1; Length 303;
Best Local Similarity 20.9%; Pred. No. 0.22;
Matches 54; Conservative 32; Mismatches 85; Indels 87; Gaps 11;
QY 78 PWYV-----ALLGFDVHGNYLV-----NKAFELTS---ITSVSILDC-----WT 115
DB 32 PHNYTYAMLLTLIIIFVGNVLVCMVAVSREKALQTTTNYLVSLAVADLLVATLVMPV 91
QY 116 IWMSIIFTWMLFGTKSYQYQVFGAAICVGGLLLVLLSDSGVTAAGSNPLLDGLVITGSI 175
DB 92 VYLEVVGEMKFSRIHCDIF-----VTL-----DVMCTASI 122

QY 176 LFTLSTVGQCYCVKRRKDRIEYVVMIGVFGMLISATEITVLERNALSSMOWSTGLLAAYV 235
Db 123 L-----NLCAISIDRYTAVAM-----PMLYNTRYSSKRRVTVMIA--IV 159
QY 236 YALSSFLCCTLPFLLKMSGAAFFNLMSLTSMDMAVAIRTFIYNQEVDMLY--YLAFCIV 293
Db 160 WLSPTISCPL-----LFGLNNTDQNECIIANPAFVYSSIVSFYVFFIVTLLV 208
QY 294 VVGIFITYKTEKDPNTR 311
Db 209 YIKIYIVLKRKRKRVNTK 226

Search completed: August 1, 2000, 19:51:36
Job time: 18758 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:31:37 ; Search time 61.65 Seconds
(without alignments)
376.755 Million cell updates/sec

Title: US-09-147-955-10
Perfect score: 1723
Sequence: 1 MSSSSRRWRENGMRRLL.....GNLDHYSLLLEDQDTPRPK 335

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL12:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	428	24.8	505	3	O59785 schizosacch
2	175	10.2	32	4	O14963 homo sapien
3	134.5	7.8	292	1	O59388 pyrococcus
4	125.5	7.3	370	3	Q04083 saccharomyc
5	122.5	7.1	341	3	O94654 schizosacch
6	120.5	7.0	411	10	P93390 nicotiana t
7	117	6.8	408	10	P92991 arabidopsis
8	114	6.6	287	2	O06933 maloncomonas
9	114	6.6	311	1	O9Y8U7 aeropyrum p
10	112.5	6.5	294	2	O9WQ07 thermotoga
11	111	6.4	225	5	O96702 iliyasaas o
12	110.5	6.4	340	10	O82600 arabidopsis
13	110.5	6.4	382	4	O9Y308 homo sapien
14	110.5	6.4	445	2	O34383 bacillus su
15	109	6.3	443	8	Q34844 katharina t
16	107.5	6.2	394	10	O24091 medicago tr
17	106	6.2	390	10	P93642 zea mays (m
18	105.5	6.1	378	2	O67513 aquifex aeo
19	105.5	6.1	434	3	O94695 schizosacch

20	105	6.1	548	8	O63543 bodo saltan
21	105	6.1	617	1	O59120 pyrococcus
22	104.5	6.1	479	2	O87134 vibrio chol
23	104.5	6.1	479	2	O87155 vibrio chol
24	104	6.0	275	1	O57793 pyrococcus
25	104	6.0	362	10	O82750 arabidopsis
26	103.5	6.0	344	2	O66073 bacillus li
27	102.5	5.9	285	1	O58347 pyrococcus
28	102.5	5.9	301	5	O19026 caenorhabdi
29	102.5	5.9	410	10	P93389 nicotiana t
30	102.5	5.9	485	2	O926M8 chlamydia p
31	102	5.9	296	1	O9YG25 aeropyrum p
32	101	5.9	273	2	O9X2A7 thermotoga
33	101	5.9	1001	11	O62439 mus musculus
34	101	5.9	1028	6	O18886 oryctolagus
35	100	5.8	328	10	O81326 arabidopsis
36	100	5.8	459	8	O21710 pygathrix r
37	100	5.8	755	8	Q9XMS0 tetrahymena
38	99.5	5.8	364	5	Q20787 caenorhabdi
39	99.5	5.8	457	8	O47876 alligator m
40	99.5	5.8	935	3	O42986 schizosacch
41	99	5.7	264	1	O28298 archaeoglob
42	99	5.7	291	2	O86204 escherichia
43	99	5.7	406	4	O9Y522 homo sapien
44	99	5.7	745	2	O924C1 escherichia
45	99	5.7	745	2	Q9WW95 escherichia

ALIGNMENTS

RESULT 1
O59785 PRELIMINARY; PRT; 505 AA.
ID O59785
DC 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE HYPOTHETICAL 56.8 KD PROTEIN.
GN SPCC320.08
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA WOOD V., RAJADREAM M.A., BARRELL B.G., WEDLER H., WAMHUTT R;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022245; CAA18310.1; -
KW Hypothetical protein.
SQ SEQUENCE 505 AA; 56780 MW; 71E0677A CRC32;

Query Match	24.8%	Score 428;	DB 3;	Length 505;
Best Local Similarity	31.48;	Pred No. 6.6e-22;		
Matches	93;	Conservative	65;	Mismatches 108; Indels 30; Gaps 7;
QY	24	LGQLVDFDLAIMTFSASIVSTVDAPLTMSTFTYVVVALLYGTILLYR-----R	72	
DB	147	LGQLVSLCITATNTENGYSINIPAFQPLVYALLTLVTPYTFRMGFKYFEMIFR	206	
QY	73	HKFLVPWYV-YALLGVDFVHGNLYLNKAFELTSTVSYLDCTIYWSIIFTWMLGTRY	131	
DB	207	HG-----WKYIIFAFDFVEGNYFVFLAIQTITNLSLSDSWATVANVILFLKRY	260	
QY	132	SVYQFVGAICVGGLLVLLSDSGVTAAGS-----NPLLDGFLVITGSILFTLSTVGQYEC	187	
DB	261	HWSQILGVVACIGGLVLLWSD--VISRGDYSANVPGLDGYMIGATCVGVSNTLEEFY	318	
QY	188	VKKRDIEVVAIGVFGMLISATETVTLERNALSSMWSN---GLLAAYVVVYALSFLFC	244	
DB	319	ASKPLPVVVGOLSLYLSIIITQITFIIDRHHLYTLHTSEMGGYLAGFI---LYNLLY	375	


```

Db 97 DNSGKLTGVLGLFAMWTLFNHYNYNKQVLK-ALHAPMTVTLVQFANGSVLI-TIM 154
QY 69 ----LYRRHKE-----LVPWYALLGVDVGHGNYLVNKAFELTSITSVILDCWTV 117
Db 155 WYLNLYRKPKISGAQLAAILP-----LAVHTLGNLFTNMSIGKVSFETHIKAMEPF 208
QY 118 WSIIFTMFLGTGYSVQFVGAAICVGLLLVLLSDGVTAAAGSNPLLDGLFVITGSLF 177
Db 209 FSVLLSANFGEKEPTPVLGAIVPIVGGVALASISVFNWAGPSANSLTNQ----- 263
QY 178 TLTSTVGQCYCKRDRLEVVAMIGVGMELISATEITVLERNALSSMOWSTGLLAAYVYA 237
Db 264 SRNVLKRVVYKDDSLDNLTFSI-----ITLMSL-----VLMAPVTF 303
QY 238 LSSFLFCTLTPELLKMSGAFFNLMLSDTMDMAVAITRTIYNOEVDWL-----YYL 288
Db 304 TEGIKF---TPSYIQSAGV---NVKQIYKSLIALCFLAY-QQVSYMLARVSPVHSV 356
QY 289 AFC-----LVVVGIFIIYKTEKDN 309
Db 357 GNCVKRVVYVSSVIFKTPSPVN 381

RESULT 8
ID O06933 PRELIMINARY; PRT; 287 AA.
AC O05933;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE MADN.
GN MADN.
OS Maltonomonas rubra.
OC Bacteria; Maltonomonas.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97274645.
RA BERG M., HILBI H., DIMROTH P.;
RT "Sequence of a gene cluster from Maltonomonas rubra encoding components of the malonate decarboxylase Na+ pump and evidence for their function.";
RL Eur. J. Biochem. 245:103-115(1997).
DR EMBL; U07980; AAC45409.1; -.
SQ SEQUENCE 287 AA; 31791 MW; 6B3209C0 CRC32;

```

```

Query Match 6.5%; Score 114; DB 2; Length 287;
Best Local Similarity 21.5%; Pred. No. 0.73;
Matches 65; Conservative 65; Mismatches 122; Indels 50; Gaps 14;

QY 32 LAIMT-----FSASLVSTVDAPLTMSFTYTVVALLYGTILLYRRHKFLVPWYVALLGP 87
Db 4 LTIKTFWAFSLSLIGVYLAGQVDAYFSLIRVSLAALVFLPFRMSVSLKIALKLMVLGG 63
QY 88 VDVHGNLY-VNKAFELTSITSVILDCWTVNSIIFTWTF--LGTKYSVQFVGAAICVG 144
Db 64 IQLCMVYFYQSFLLTVPEVLI---FPIITPIYVTLFIDLMQCFSHRYLLTAYLAVV 120
QY 145 GLLLVLLSDGVTAAAGSNPLLDGLFVITGSLFSTVGQCYCKRDRLEVVAMIGVFG 204
Db 121 GAFIRYSN-----VGDNLVAGFVVGQAGNICFALGQGVYKLLIER-EKLE-LSORSVFG 173
QY 205 -----MLISATEITVL---ERNALSSMOWSTGLLAAYVYVYALSFLFCTLTPELLKMSG 255
Db 174 YFYLGAIIATAWAIFGNEKFLPATSLQWGLVYLVGVIASGIGYFLN-----NKG 224
QY 256 A-----AFFNLMLSDTMDMAVAITRTIYNOEVDWL--YYLACLVVVGFIYIT---KT 303
Db 225 ATKVDVTGVLAIMNNALIPA---GLIVNLLIWNRDADVLRLFTGGSIIALSILHFEFWGKT 281
QY 304 EK 305

```

```

Db 282 KK 283

RESULT 9
ID Q9Y807 PRELIMINARY; PRT; 311 AA.
AC Q9Y807;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 311AA LONG HYPOTHETICAL PROTEIN.
GN APE2537.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K1.
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000064; BAA81553.1; -.
SQ SEQUENCE 311 AA; 32703 MW; 15345528 CRC32;

Query Match 6.6%; Score 114; DB 1; Length 311;
Best Local Similarity 20.3%; Pred. No. 0.79;
Matches 62; Conservative 51; Mismatches 127; Indels 66; Gaps 11;

QY 32 LAIMTFSASLVSTTV-----DAPLTMSFTYTVVALLYGTILLYRRHKFLVPWY 82
Db 9 LALAVANVSTASILVRLAGVGFAGAAATWRLVLTILTALVAAYAVKRRKVP-- 64
QY 83 ALLGFPD-----VHGNYLVNKAFELTSITSVILDCWTVNSIIFTWTF 130
Db 65 ---GGVDLLVLAASGVALLHFDLWASLFHMSVMTVVDSPYAVLAIVGFRLF-KEK 120
QY 131 YSVQFVGAAICVGGLL-LVLLSDGVTGTA-AGSNPLLDGLFVITGSLFSTVGQCYC 188
Db 121 YTPLQLLGAVIAMVGVAGIAYFYSRGASAPPGDPIKGFLLAFAGML-----AVAAVFSI 175
QY 189 -----KRRDRIEVAMIGVGMELISATEITVLERNALSSMOWSTGLLAAYVYVYALSFLF 243
Db 176 GKGLRSKYSTILEYTLVYVGAIAIVSLLTTLVVRAPL-----TG-----YTAETLY 222
QY 244 CYLTPFLKMSGAFFNL-----SMUTSDMNAVAITRTIYNOEVDWLYYLAFC 291
Db 223 LITLALLPMLGHTLILNLLRLRLSLAATVPILGEPVGAALAWILLGKYSTLEALLMV 282
QY 292 LVVVG 297
Db 283 VYLAGI 288

RESULT 10
ID Q9WYQ0 PRELIMINARY; PRT; 294 AA.
AC Q9WYQ0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE SUGAR ABC TRANSPORTER, PERMEASE PROTEIN.
GN TM0419.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
RN [1]
RP SEQUENCE FROM N.A.

```


QY 221 SSMQSTGLAAYVYVALSSFLCTLPFLKNSGAAPFNLMLTSDMNAVAIRFIYNO 280
 Db 200 FLMPDLTAIDANTILIYAM-----GQSFLLSVGVSM--VTYSAYLEKQ 241
 QY 281 E 281
 Db 242 E 242

RESULT 15

Q34844 PRELIMINARY; PRT; 443 AA.
 AC Q34844;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE NADH DEHYDOGENASE SUBUNIT 4.
 OS Katharina tunicata (black chiton).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Mollusca; Polyplacophora; Neoloricata;
 OC Ischnochitonida; Mopaliidae; Katharina.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95129806.
 RA BOORE J.L., BROWN W.M.;
 RT "Complete DNA sequence of the mitochondrial genome of the black
 chiton, Katharina tunicata.";
 RL Genetics 138:423-443(1994).
 DR EMBL; U09810; AAC48367.1; -.
 DR PFAM; PF00361; oxidored_q1; 1.
 DR PFAM; PF01059; oxidored_q5_N; 1.
 KW Mitochondrion.
 SQ SEQUENCE 443 AA; 50447 MW; 4927BF1C CRC32;

Query Match 6.3%; Score 109; DB 8; Length 443;
 Best Local Similarity 20.2%; Pred. No. 2.5;
 Matches 68; Conservative 61; Mismatches 109; Indels 98; Gaps 18;
 QY 30 FDLAIMTFSASLVSTVDAPLTMSTTTVVVALLYGTILYRRHKFLVPMY--YALLG- 86
 Db 105 FSLYIF-FEASLIPT-----LIILMWG----YQPERLOAGMYMIYTLGA 146
 QY 87 ---FVDVGHNYLVNKAFLTSTTSVSLDCWTIVMSIIFTWMFL---GTKYSVYOF--- 136
 Db 147 LPPLINIFFIYSHNAHLNLLIMSLPIMPYQAM---ISFWMLFIILVFLVKLPYISFHLW 203
 QY 137 -----VGAICVGGLLVLVLSDSGVTRAGSNP-----LIGDFLVTGSLFTLSTV 182
 Db 204 LPAHVEAPVAGSMILAAALL-KGGYGINTINNIFKNFVNLVFIITLGLVGGVLSL 262
 QY 183 GQBYCYKVRKDRIEWA--MIGVFGM-----LISATEI----- 212
 Db 263 ---ICLRQSDMKALTAYSVGHMGLMAGLLSFEGLKWMALLMTAHLSSGLFSMSN 319
 QY 213 TVLERALSSMQNSQGLAAYVYVALSSFLCTLPFLKNSGAAPFNL-----SMLTSDM 268
 Db 320 MMYKSGSRLFITGFIISLVPTFSNFMCS-----INMAAPPSINLLSEAGLMIS--- 372
 QY 269 WAVAIRFIYNOEVDWLYLAFCLVVGVIYIYKTE 304
 Db 373 -TVSISKFL----MILLALMAFVSAVYTLFLFTNTQ 403

Search completed: August 1, 2000, 21:31:42
 Job time: 6071 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:32:20 ; Search time 31.57 Seconds
(without alignments)
328.836 Million cell updates/sec

Title: US-09-147-955-10
Perfect score: 1723
Sequence: 1 MSSSSRRWRNENGRMTLL.....GNLDHEYSLLDQDDTPRKP 335

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145	8.4	324	1 YAGC_SCHPO	Q09875 schizosacch
2	128	7.4	304	1 PAGO_SALTY	O30846 salmonella
3	124.5	7.2	305	1 YBVB_BACSU	O32256 bacillus su
4	124	7.2	330	1 YETK_BACSU	O31540 bacillus su
5	121	7.0	306	1 YEDA_ECOLI	P09185 escherichia
6	119	6.9	490	1 CAR3_DICDI	P35352 dictyosteli
7	109.5	6.4	541	1 YD18_MYCTU	Q10631 mycobacteri
8	108.5	6.3	330	1 Y355_SYNT3	P74436 synechocyst
9	106.5	6.2	531	1 PTM1_YEAST	P32857 saccharomyc
10	106	6.2	402	1 CPT2_BRAOL	P52178 brassica ol
11	105	6.1	306	1 YBFH_BACSU	O31448 bacillus su
12	105	6.1	549	1 COX1_TRYBB	P04371 trypanosoma
13	105	6.1	1021	1 TSCC_HUMAN	P55017 homo sapien
14	103.5	6.0	311	1 YCAY_CLOAKL	P38943 clostridium
15	103.5	6.0	393	1 YMB8_YEAST	Q03730 saccharomyc
16	103.5	6.0	637	1 YHE7_YEAST	P38724 saccharomyc
17	102.5	5.9	686	1 NU5M_SCHCO	P50368 schizophyll
18	102	5.9	411	1 PR23_RABIT	P46069 oryctolagus
19	102	5.9	517	1 QOK1_SULAC	P98004 sulfolobus
20	101.5	5.9	312	1 OLF2_CHICK	P37068 gallus gall
21	101	5.9	459	1 NU4M_PELSU	O79677 pelomedusa
22	100.5	5.8	318	1 OLF1_CHICK	P37067 gallus gall
23	99.5	5.8	461	1 PUCC_RHOCA	P23462 rhodobacter
24	99.5	5.8	970	1 Y277_MYCGE	Q49409 mycoplasma
25	99	5.7	432	1 Y672_MERJA	Q58086 methanococc
26	99	5.7	474	1 NU4M_DIDMA	P41308 didelphis m
27	99	5.7	1105	1 YEGE_ECOLI	P38097 escherichia
28	98.5	5.7	339	1 NUOH_RICPR	O92cf7 rickettsia
29	98.5	5.7	444	1 CADB_ECOLI	P23891 escherichia
30	98.5	5.7	488	1 LYSP_ECOLI	P25737 escherichia
31	98	5.7	301	1 YIJE_ECOLI	P32667 escherichia
32	97.5	5.7	407	1 GLUP_HELPY	O25788 helicobacte
33	97.5	5.7	409	1 NU4M_CAEEL	P24892 caenorhabdi

34	97.5	5.7	572	1 NU5M_LOCOMI	Q36428 locusta mig
35	97	5.6	273	1 BACA_ECOLI	P31054 escherichia
36	97	5.6	343	1 Y841_METTH	O26929 methanobact
37	96	5.6	326	1 UGNT_CANFA	O77592 canis famil
38	96	5.6	574	1 NU5M_DROME	P18932 drosophila
39	95.5	5.5	340	1 COX1_CRION	P98003 crithidia o
40	95.5	5.5	444	1 D2DR_BOVIN	P20288 bos taurus
41	95.5	5.5	448	1 PCAK_PSEPU	Q51955 pseudomonas
42	95.5	5.5	609	1 NAH_CAEEL	P35449 caenorhabdi
43	95.5	5.5	893	1 PMTX_SCHPO	O13898 schizosacch
44	95	5.5	289	1 Y510_ARCFU	O29740 archaeoglob
45	95	5.5	580	1 NU5M_ANOGA	P34854 anopheles g

ALIGNMENTS

RESULT	1
YAGC_SCHPO	
ID	YAGC_SCHPO STANDARD: PRT; 324 AA.
AC	Q09875;
DT	01-FEB-1996 (Rel. 33, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-FEB-1996 (Rel. 33, Last annotation update)
DE	HYPOTHETICAL 35.8 KDA PROTEIN C12G12.12 IN CHROMOSOME I.
GN	SPAC12G12.12.
OS	Schizosaccharomyces pombe (Fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC	Schizosaccharomycetaceae; Schizosaccharomyces.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-972;
RA	Devlin K., Odell C., Churcher C.M., Barrell B.G., Rajandream M.A.,
RA	Walsh S.V.;
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	EMBL; 266568; CAA91507.1;
DR	Hypothetical protein; Transmembrane.
KW	TRANSMEM 34 54 POTENTIAL.
FT	TRANSMEM 76 96 POTENTIAL.
FT	TRANSMEM 103 123 POTENTIAL.
FT	TRANSMEM 127 147 POTENTIAL.
FT	TRANSMEM 158 178 POTENTIAL.
FT	TRANSMEM 198 218 POTENTIAL.
FT	TRANSMEM 243 263 POTENTIAL.
FT	TRANSMEM 275 295 POTENTIAL.
FT	TRANSMEM 297 317 POTENTIAL.
SEQUENCE	324 AA; 35746 MW; 9696681A0E82CCD3 CRC64;

Query Match 8.4%; Score 145; DB 1; Length 324;
Best Local Similarity 23.8%; Pred. No. 0.00087;
Matches 67; Conservative 42; Mismatches 117; Indels 56; Gaps 10;

QY	68	LLYRRHKFLVPWY-----WYALLGFVDVHGNYLNKAFETLSITSVSILDCWT	115
Db	53	VLKRRHS-QQPGVESLDHPLKHKVFMALPAIDICGSLMNVGLLTSIYQMTFGSL	111
QY	116	IYMSIITFMFLGTRYVYQFYGAACVCGLLLLVLLSDSGVTAAGSNPLGLDFLIVTGS	175
Db	112	IIIFALFATPLLKRTIQQLQWLSLGFVVLGVAIVGSGSS-SSIGSNPILGTAIVLIGQF	170
QY	176	LFILSTVGQEVCK--RKDRIFVAMIGVFGMLISATETITVLERNALSSMOWSTGLAAY	233

Db 171 FLATQFTIEEVLFIQVDPSELVAYEGTVGVFFLLGMII-----SYFIGSTTACY 223
 QY 234 -----VVVALSSEFLCTLTPELLKMSGA-----AFFNLSML-----T 265
 Db 224 HGWFDYSHVISRF---NEVPALYVISCIVIVSVIAFFNVSGLAITKLHSAATTRSLDIART 280
 QY 266 SDMNAVAIRTIYNOEVDWLYLAFLCVVGVGIYTKTKDP 307
 Db 281 FGINIINAMCH--ESPHLOFGLGVLLIYGIYTHSIKFP 320

RESULT 2
 ID PAGO_SALTY STANDARD; PRT; 304 AA.
 AC Q30646;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE PAGO PROTEIN.
 GN PAGO.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NTCC 14028S;
 RA Gunn J.S., Belden W.J., Miller S.I.;
 RT "Identification of Phop-PhoQ activated genes within a duplicated
 region of the Salmonella typhimurium chromosome.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF013775; AAB82452.1; -
 DR STYGENE; SG27777; PAGO.
 DR PFAM: PF00892; DUF6; 2.
 KW Transmembrane.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 34 54 POTENTIAL.
 FT TRANSMEM 67 87 POTENTIAL.
 FT TRANSMEM 95 115 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 246 266 POTENTIAL.
 FT TRANSMEM 267 287 POTENTIAL.
 SQ SEQUENCE 304 AA; 33622 MW; B368153FD6BFB0B CRC64;

Query Match 7.4%; Score 128; DB 1; Length 304;
 Best Local Similarity 21.9%; Pred. No. 0.013;
 Matches 61; Conservative 53; Mismatches 119; Indels 46; Gaps 10;
 QY 59 VVALYGTILLY----RRHKFLVPWYALLGFVDVHGNLYNKAFFELTSITSVILDCW 114
 Db 50 IIAWLRKKTLLFPFGQRLFOFVICIFVFCIPFSLMYGETYVNSGLAAIIFANMPVA--- 106
 QY 115 TIWWSIIFTWFLCTKYSVYOFVGAATCVGGLLLVLSDS-----GYTAAGSNPLL 165
 Db 107 -----VLIASVFLNEKAKMOIAGLTALTALTGILLETNTSTSHQGITALLIS----- 158
 QY 166 GDFLVITSGILFTLSTVQGEYCVKRDRIEYVAMIG-----VFGMLISATEITVLNALS 221
 Db 159 ---AVLIHAIIT-----QCKRSCVSVITFNALPCLLAGLILSATG-WFFERPQVS 207

QY 222 SMOGSTGLAAVYVVALSSP--LFTTLTPPELLKMSGAAP-FNLISMLTSDMNAVAIRTIY 278
 Db 208 TFSVSHILATLY----LGAFAVGFIILCYFALOOKANAFQASLVFLIFLPIAIVSLEDIY 263
 QY 279 NOEVDWLYLAFLCVVGVGIYTKTKDPNNTALENGN 317
 Db 264 GYAISTHSMILLIIPLVIGIFLTVLARNLPVTSRCRNSS 302

RESULT 3
 ID YVBV_BACSU STANDARD; PRT; 305 AA.
 AC Q32256;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL 32.8 KDA PROTEIN IN ARAR-LACA INTERGENIC REGION.
 GN YVBV.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z99121; CAB15405.1; -
 DR SUBTILIST; BGL4088; YVBV.
 DR PFAM: PF00892; DUF6; 2.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 11 31 POTENTIAL.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 97 117 POTENTIAL.
 FT TRANSMEM 126 146 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 244 264 POTENTIAL.
 FT TRANSMEM 265 285 POTENTIAL.
 SQ SEQUENCE 305 AA; 32820 MW; 203B2E4E7138670B CRC64;

Query Match 7.2%; Score 124.5; DB 1; Length 305;
 Best Local Similarity 23.3%; Pred. No. 0.024;
 Matches 74; Conservative 51; Mismatches 131; Indels 61; Gaps 13;
 QY 27 LVSPDLATM--TTSASLVSTVTDAPLPTMSFTTYVVALLYGTLIRRHKLVPWY-- 81
 Db 12 LLAFLVIMWGVNPNPLSKAALAYSPPLIFAGIRTLGGLLLVVALPRIHLRKETWPI 70
 QY 82 ---VALLGFVDVHG-----NVLNKAFFELTSITSVILDCWTIYWSIIFTWMLGTKYS 132
 Db 71 YLVSAALLNTLFYGLTIGLNYLPAGLF-----SAIVFPQVLMGVFSLMLGESMF 122
 QY 133 YVQFGAAICVGLLLLVLLSDSGVTAGSNPLLGDFLITGLTSTVGOEYCVKRD 192
 Db 123 VMKVIGLITGFAGVAVISAAGFG---GHISVIGLVLLGSAVSWALGTYYVKKTKGSVD 178
 QY 193 RIEVAM---IGVFGMLISA--TEITVLNALSMMOWSTGLLAAYVVVYVLSLFLCTLT 247

```

Db 179 SIWVALQITGVSFLLISGPWE-----SPSAIOWTAPF-----ITSLLF--IS 221
QY 248 PELLKMSGAARFNLSMLTSDMWAVAITF-----IYNQVDWMLYYLAFCLVVG 296
Db 222 VVIALGLWLVFTL-VSGGEASKVASYFLIPLISIVASSIFLHELPLTSLLAGLLIYT 280
QY 297 IFIYTKTEKDPNTRAL 313
Db 281 SICLVNTRSKAKAAAI 297

RESULT 4
ID YETK_BACSU STANDARD; PRT; 330 AA.
AC O31540;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 35.8 KDA PROTEIN IN LFED-PEL INTERGENIC REGION.
GN YETK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z99107; CAB12540.1; .
CC DR SUBTILIST; BG12867; YETK.
CC DR PFAM: PF00892; DUF6; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 56 76 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
SQ SEQUENCE 330 AA; 35758 MW; 7C6C65F1EDB276B5 CRC64;

Query Match
Best Local Similarity 20.78; Pred. No. 0.028;
Matches 66; Conservative 48; Mismatches 139; Indels 66; Gaps 9;

QY 20 LGLGLGQLVSDLAIMTFS-----SLVSTTVDAPLTMSFTTYV--VALLYGTHLLYR 71
Db 58 LSSGLRFLIASVILMLLFCLEKGFPAITKDVFLVQSGFTGVFLSICLLYGV----- 112
QY 72 RHKLVPVWYVALLGFVDVHGNLVNKAFLTSTVSILDCWTIVWSIITFWMLGKY 131
Db 113 -----QTTGTSGILSTTPTMLIGILSFFLLREKI 143
QY 132 SVYQFVGAICVGGILLVLLSDSGVTAAGSNPLLGDFLVIT---GSIFFLTSGVQECV 188
Db 144 EKKTLIGILLACVGMALNFGAGSDGTGPHALFGNMLIIAAVIGALFTLMA---KLLS 200
QY 189 KRKRIEVVAMIGVGM-----ISATETIVLERNALSSMQWSTGLLAAVYVALSSFLFC 244

```

```

Db 201 PHISALAIATVSVLSGFLFFLPALFPAASSFDYSVPVFLDWS-----YVLYIA---LFV 251
QY 245 TLTPFLKMSG-----AAAFNLNLSMLTSDMWAVAITFYNQVDWMLYYLAFCLVVGIF 299
Db 252 TVLAFVLYWYSGVTKPACVSGIFTSVLPVSAVLISGVILKEPFVHFVFGIACVIGIRV 311
QY 300 YTKTEKDPNTRALEGNL 318
Db 312 TVIKKQPDPAEAEKTL 330

RESULT 5
ID YEDA_ECOLI STANDARD; PRT; 306 AA.
AC P09185;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 32.2 KDA PROTEIN IN DSRA-VSR INTERGENIC REGION (ORF 4).
GN YEDA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 89345179.
RA Hancik T., Gerwin N., Fritz H.-J.;
RT "Nucleotide sequence of the dcn locus of Escherichia coli K12.";
RL Nucleic Acids Res. 17:5844-5844(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97251358.
RA Itoh T., Aiba H., Kasaai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Isono K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X13330; CAA31708.1; .
CC DR EMBL; AE0000287; AAC75025.1; .
CC DR EMBL; D90835; CAB21710.1; .
CC DR PIR; JS0266; JS0266.
CC DR ECOGENE; EG11141; YEDA.
CC PFAM; PF00892; DUF6; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 7 27 POTENTIAL.

```

```

FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
SQ SEQUENCE 306 AA; 32194 MW; 6F0D63E8090ACA64 CRC64;

Query Match 7.0%; Score 121; DB 1; Length 306;
Best Local Similarity 23.4%; Pred. No. 0.042;
Matches 62; Conservative 54; Mismatches 121; Indels 28; Gaps 10;

QY 49 PLTSTFTYVVALYCTILYRHRKF--LVPWYVALLGFDVH-GNYLYNKAFELTSTI 105
DB 36 PLMAGVRFLLAAGLILLAFLLRHLPLRLLNALIGLLLAAGNGWTVVA-EHONY 94
QY 106 TS-----VSILDCWTIVNSIIFTMFLGKYSVYQVGAATCVGGLLVLLSDSGVTAA 159
DB 95 PSGLAAVVAATVPLFTLCFSLP-----GIKTRKLEWVGIAIGLAGIIML---NSGGNLS 146
QY 160 GSNPLLGDELIVNGSILFTLSTVQECYVKRKORLEVAMGVGFMILISATEITVLNERNA 219
DB 147 G-NP-WGAILILIGTSWAGSV-----YGRITLPVGMAGAEIMLAAGVVLMIASMA 199
QY 220 ---LSSMOWSTGLLAAYVVALYSSFLCTLTPFLKMGSAFFNLMLTSDMNAVAIRTF 276
DB 200 GEKLTALPSLGSFLAVGLALGFSIIAINAYMTLRNVPALATSYAVYVNPVAVLLGTG 259
QY 277 IYNGEVDWLYLAFCLVWVGIFITY 301
DB 260 LGGTSLKIEWLALGVIVFANLVLT 284

```

```

RESULT 6
CAR3_DICDI STANDARD; PRT; 490 AA.
ID CAR3_DICDI
AC P35352;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYCLIC AMP RECEPTOR 3.
GN CARC OR CAR3.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryote; Dictyostelida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE; 93170666.
RA Johnson R.L., Saxe C.L. III, Gollop R., Kimmel A.R., Devreotes P.N.;
RT "Identification and targeted gene disruption of car3, a G-protein coupled receptor subtype expressed during multicellular stages of Dictyostelium development."
RT Dictyostelium discoideum.
RL Genes Dev. 7:273-282(1993).
CC -1- FUNCTION: RECEPTOR FOR CAMP. COORDINATES THE AGGREGATION OF INDIVIDUAL CELLS INTO A MULTICELLULAR ORGANISM AND REGULATES THE EXPRESSION OF A LARGE NUMBER OF DEVELOPMENTALLY REGULATED GENES. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: INDUCED AT EARLY AGGREGATION (6 HRS) AND MAXIMALLY EXPRESSED AT THE MOUND STAGE (9-12HRS), LEVEL OF EXPRESSION PEAKS AGAIN DURING THE SLUG STAGE (18HRS) AND DECLINES AT CULMINATION.
CC -1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 5 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-15 IS THE INITIATOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S55235; AAB25437.1; -
DR PIR; A46391; A46391.
DR GCRDB; GCR_0276; -
DR DICTYDB; DD02033; CARC.
DR PRINTS; PR00247; GPCRAMP.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Multigene family.
FT DOMAIN 1 23 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 24 43 1 (POTENTIAL).
FT DOMAIN 44 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 2 (POTENTIAL).
FT DOMAIN 79 94 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 95 120 3 (POTENTIAL).
FT DOMAIN 121 131 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 132 150 4 (POTENTIAL).
FT DOMAIN 151 173 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 174 192 5 (POTENTIAL).
FT DOMAIN 193 216 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 217 235 6 (POTENTIAL).
FT DOMAIN 236 246 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 247 271 7 (POTENTIAL).
FT DOMAIN 272 290 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 204 204 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT DOMAIN 303 439 ASN-RICH.
FT DOMAIN 399 427 POLY-ASN.
SQ SEQUENCE 490 AA; 56161 MW; A28BA834D8626153 CRC64;

```

```

Query Match 6.9%; Score 119; DB 1; Length 490;
Best Local Similarity 22.3%; Pred. No. 0.094;
Matches 79; Conservative 47; Mismatches 116; Indels 112; Gaps 21;

QY 18 TLLGIG---LQQLVSFDLAIMTF--SASLVSTVTDAPLTM-----SFTTYTWALLY 64
DB 39 TLVLGFWRLKLLRNHITKIITFCSTSLAKDLITLTLEKQSGSFQCY-----LY 93
QY 65 GTILLYRRHFLPYWYVALLGFDVHGVNLYNKAFELTSTVSILDCWTI--VNSIIF 122
DB 94 ATVIYGS---LACWLTLCLSP-SIY-NLIVKREPPEKPKYHYFVFWVPFMSVIM 148
QY 123 -----TWMFLGTSYVQY---FVGAATCVGGLLVLLSDSGVTAAGSNPLLGDEL 169
DB 149 LSKGVIEVTGNMCWIGNTVYVGRFLGYGFLAIFLAAVLV---GLTSRYT-----YK 199
QY 170 VITGSILTLSTVQECYVKRDR-----IEVYAMGVGFMILISATEITVLER--NALSS 222
DB 200 VIRSV-----SDNKRHMTYQFKLINTIIVP---LLCWFAVINRIVNGLNM 244
QY 223 MOWSTGLLAAYVVALYSSFLCTLTPFLKMGSAFFNLMLTSDMNAVAIRTFIYQEV 282
DB 245 FPAWVSILHTY-----LSVSHGFVASV-----TEIYNPL 274
QY 283 DWLYLAFCLV---VVGIFITY--KTEKDPNN-----PRALENGHLDHEYS 323
DB 275 MRYLASILLIPFTKFGYFVETQORLEKNKNNNHSPVGLSNNAQNHHHHN 328

RESULT 7
YD18_MYCTU STANDARD; PRT; 541 AA.
ID YD18_MYCTU
AC Q10631;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 59.3 KDA PROTEIN RV1318C.
GN RV1318C OR MTCY130.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:H37RV;
 RX MEDLINE; 98295987.
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gencies S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultston J.E.,
 RA Taylor K., Whitehead S., Barrett B.G.;
 RT Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.*;
 RL Nature 393:537-544(1998).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-3 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 273902; CAA98105.1; -;
 DR TUBERCULIST; RV1318C; -;
 DR PFAM; PF00672; DUF5; 1;
 DR PFAM; PF00211; guanylate_cyc; 2;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 SQ SEQUENCE 541 AA; 59331 MW; 24949CC8214386B1 CRC64;

Query Match 6.4%; Score 109.5; DB 1; Length 541;
 Best Local Similarity 22.0%; Pred. No. 0.49; Indels 121; Gaps 15;
 Matches 67; Conservative 43; Mismatches 74;

QY 3 SSSRRRRENEGRRT-----LGLGLGVLSFDLAIMTFSASLVSTTVDAPLMTSFT 55
 DB 40 SESQRRRVRIQVMTALVYVNTANLLGIGV-----LLVYIALPEPSIVRTPRWLTG 93
 QY 56 T---YVVVALLGYTLLR-----RHKFLVW-----YW--- 81
 DB 94 VPGVYLLALAGLSVALRQTVQALRWAEGRKPTRBEERTFLAPRWAVGHLMFPGVG 153
 QY 82 YALLGFVDVHGNYLVNKAFLTSITSILDCWTVIVWSILFTWMLGTYSVYQVFGAAI 141
 DB 154 TALL--TLYG--LNNAFIPREFAVSF--CGVLV---ATATYLHTEALRPPFAAQL 203
 QY 142 -----CVGGLLLVLLSDSGVTAAGSNPLLDGFLVITGSLFSLSTVQGYCVK 189
 DB 204 EAGPPRRAPILGRTVMVLLGSGV-----PVVG----- 234
 QY 190 RKDRIVAMIGVGNLISATEITVLERNALSSMOWSTGLAAVYVVALSSFL-----F 243
 DB 235 -----IALMAFMFVLL-----NLTRMQFATGVLLISMVTLVFGFILMWILAW 277
 QY 244 CTLP 248
 DB 278 LTATP 282

RESULT 8
 Y355_SNN3
 ID Y355_SNN3 STANDARD; PRT; 330 AA.

P74436;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL 36.1 KDA PROTEIN SLL0355.
 GN SLL0355.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97601201.
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugliara M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraiki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
 RT genome and assignment of potential protein-coding regions.*;
 RL DNA Res. 3:109-136(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D90915; BAA18537.1; -;
 DR PFAM; PF00892; DUF6; 2;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 41 61 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 102 122 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT TRANSMEM 201 221 POTENTIAL.
 FT TRANSMEM 238 258 POTENTIAL.
 FT TRANSMEM 284 306 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 SQ SEQUENCE 330 AA; 36092 MW; AB44A99D8B53DAC CRC64;

Query Match 6.3%; Score 108.5; DB 1; Length 330;
 Best Local Similarity 21.5%; Pred. No. 0.36; Indels 59; Gaps 15;
 Matches 71; Conservative 59; Mismatches 141;

QY 13 EGMRTLLGLGLGVLSFDL---AINTFSASLVSTTVDAPLMTSFTYTVVALLYGTILL 69
 DB 4 ESKTNTNIRSGTLTAPLPEFLMGTAAMKGVLDATTPFFVATVRLIPAGILVLLWANGOK 63
 QY 70 YRHKFLVPMYVALLGVVD--VHGNYLVNKAFLT--SITSVSLDCWTVIVWSILFTW 125
 DB 64 RPOQWQWQWMIILFALVDGTLFGFLA--OGLERTGAGLGSV--IIDSQIALLSSNL 121
 QY 126 F-----LGTKYSYQVFGAAICVGGLLLVLLSD-----SGVTAAGSNP 163
 DB 122 FREVIGGIG-----WLLGLLVGGISLIGLPDPEWFLQHLGLQGLSINMSG--SALGSS- 172
 QY 164 LLGDFLVITGSLTSLTVGOEYCVKRDRIEVVAMIGVFGMLISATEITVLERNALSSM 223
 DB 173 --GELMLLASLSMAGVGLVLPFVSRVDPVAVTGWMIIGGLPLLAIALVQDSERQNI 230
 QY 224 Q-WSTGLLAAYVYV--ALSSFLCTLTPFLKMSGAAFFNLSMLTS-----DMWVAIRT 275
 DB 231 DWGNGNLAYATVFGSATAYGIF-----FYLSKSG----NLTSLSLSTFLTPFALSFSN 281
 QY 276 FIYNQEDMLYIACLVVVGIFVIYTKTEK 305

DB 282 LILBQLSSLOWLGVAFITLYIYLINOREQ 311

```

RESULT 9
ID PTM1_YEAST STANDARD; PRT; 531 AA.
AC P32857;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE PROTEIN PTM1 PRECURSOR.
GN PTM1 OR YKL039W OR YKL252.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RN SEQUENCE FROM N.A.
RA Hearer B.K., Petzold A.S., Brown S.S.;
RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 93127731.
RA Purnelle B., Skala J., van Dyck L., Goffeau A.;
RT "The sequence of a 12 kb fragment on the left arm of yeast chromosome
RT XI reveals five new open reading frames, including a zinc finger
RT protein and a homolog of the UDP-glucose pyrophosphorylase from
RT potato."
RL Yeast 8:977-986(1992).
CC -!- SIMILARITY: STRONG, TO YEAST YHL017W.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11895; AAA34921.1; -
DR EMBL; X69584; CAA45300.1; -
DR EMBL; Z28039; CAA81874.1; -
DR PIR; S30011; S30011.
DR SGB; L0001524; PTM1.
KW Transmembrane; Signal.
FT SIGNAL 1 ? POTENTIAL.
FT CHAIN ? 531 PROTEIN PTM1.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 230 247 POTENTIAL.
FT TRANSMEM 270 294 POTENTIAL.
FT TRANSMEM 305 325 POTENTIAL.
FT TRANSMEM 334 354 POTENTIAL.
FT TRANSMEM 379 402 POTENTIAL.
FT CONFLICT 513 531 ATGQCPLQONNATSSRPD -> GHGPVSPSPK (IN
FT REF. 1).
SQ SEQUENCE 531 AA; 61006 MW; 82BE7D22878E12AD CRC64;

Query Match 6.2%; Score 106.5; DB 1; Length 531;
Best Local Similarity 21.7%; Pred. No. 0.79;
Matches 70; Conservative 48; Mismatches 107; Indels 97; Gaps 18;

QY 57 YTVVALLYGTYLLYRRH-----KFLVPVWYALLGFVDVHGNY-LVNKAFELTSI----- 105
DB 206 YVAMALYSFAFKKHKHLLPLQKYLALFAFFVLTAEITFVWAYDYLKNEKGDGTAGIKVYM 265
QY 106 TSVILDCWTIVWSIIFTNWFLGKTSVY-----QFVGA---AICVGLLLVL 150
DB 266 VFLSILTAGKVTFES-PELLLIITAGLYGIVPKLTKLRRRCOMGYAICIGFLQSY 324
QY 151 LSDSGVTAAGSNPLGLDFLVTISILFTLSITGQBYCYVKRKDRIEVAMIGVFNLSAT 210
DB 325 LTD----NEAPSPL-----LITLIPM-----ALALIIFYMYIIRSM 357

```

```

QY 211 EITVL-----ERNALSSQWSTGLLAAYVYVALSSFLCTLTPELLKMSG----AAFENLSML 264
DB 358 TKTVYLKEQRIQVKNMYKKLL--YIIYA-----SFLSVLAGSIYSSFIYVGMN 405
QY 265 TSDM-----WAVAIRTFIYNOEVDWLYLAFCLVWVGIFYTYKTE-----KDPN 308
DB 406 TIDMEKNW--RSREFFYDFMPTLVYFIVFVITA---FLMRPTDTSYMLAASQQLPTDPE 460
QY 309 NTRALENGNLDHEYSLLDQDD 330
DB 461 NVADFGLDGLQS-----FDDQDD 478

RESULT 10
CPT2_BRAOL STANDARD; PRT; 402 AA.
ID P52178;
AC P52178;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRIOSE PHOSPHATE/PHOSPHATE TRANSLATOR, NON-GREEN PLASTID PRECURSOR
DE (CTPT).
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Brassica.
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. BOTRYTIS; TISSUE=FLOWER;
RX MEDLINE; 97246375.
RA Fischer K., Kammerer B., Gutensohn M., Arbinger B., Weber A.,
RA Hausler R.E., Fluegge U.-I.;
RT "A new class of plastidic phosphate translocators: a putative link
RT between primary and secondary metabolism by the
RT phosphoenolpyruvate/phosphate antiporter."
RT Plant Cell 9:453-462(1997).
CC -!- FUNCTION: MEDIATES THE EXPORT OF FIXED CARBONS FROM THE
CC CHLOROPLASTS INTO THE CYTOSOL IN THE FORM OF TRIOSE PHOSPHATES.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LOCATED IN
CC ZONES OF CONTACT BETWEEN THE INNER AND OUTER MEMBRANE OF THE
CC CHLOROPLAST (BY SIMILARITY).
CC -!- SIMILARITY: STRONG, TO OTHER PLANTS CTPT. ALSO SIMILAR TO
CC YEAST PROTEIN SLY41.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13632; AAA84892.1; -
KW Transmembrane; Chloroplast; Transit peptide; Transposon.
FT TRANSIT 1 82 CHLOROPLAST (POTENTIAL).
FT CHAIN 83 402 TRIOSE PHOSPHATE/PHOSPHATE TRANSLATOR,
FT NON-GREEN PLASTID.
FT DOMAIN 83 99 INTERMEMBRANE SPACE (POTENTIAL).
FT TRANSMEM 99 119 POTENTIAL.
FT DOMAIN 120 131 LUMENAL (POTENTIAL).
FT TRANSMEM 132 152 POTENTIAL.
FT DOMAIN 153 209 INTERMEMBRANE SPACE (POTENTIAL).
FT TRANSMEM 210 230 POTENTIAL.
FT DOMAIN 231 278 LUMENAL (POTENTIAL).
FT TRANSMEM 279 298 POTENTIAL.
FT DOMAIN 299 374 INTERMEMBRANE SPACE (POTENTIAL).
FT TRANSMEM 375 394 POTENTIAL.
FT DOMAIN 395 402 LUMENAL (POTENTIAL).
FT SEQUENCE 402 AA; 43523 MW; 4970CCFF6E77CB03 CRC64;
SQ

```


FT	TRANSMEM	248	268	POTENTIAL.
FT	TRANSMEM	271	291	POTENTIAL.
SEQ	SEQUENCE	306 AA;	33842 MW;	7DE953E93B2FB072 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 306;
 Best Local Similarity 21.2%; Pred. No. 0.59;
 Matches 64; Conservative 63; Mismatches 143; Indels 32; Gaps 14;

QY	32	LAIMFESASLSTVDADLTMSFTTYIVV--ALLYGTILLYRRHFLVPW-YW-----YA	83
DB	15	VTLLMGTTFTSTKV---LLADFGPMELPYFLMFGIALILVPRNMIPFRNMQELLFA	71
QY	84	LLGFVDVHGNYLV-NKAFELSTSVS-ILDCWTIVMSIIFTMFLGTXYSVYQVGA	141
DB	72	GAGLFGVTYFLLENIALTYVASNVGMIVSIIPMITAVLAHFLLEGEKRLTFLG	131
QY	142	CVGGLLIVLLSDGVTAGSNPL-----LGDFLVIVGSLFTLSTVGQYCVKRDREV	196
DB	132	ALIGLLLTFT--NGNVURLPLDGMIAAGAAVFGVGYSTFMKKLSAYEYHI-----IEL	184
QY	197	VAMIGVFGMLISATITVLERN-ALSSWQSTGLL-AAVYVALSSFLFCTLTPELLKMS	254
DB	185	TQRFVLYGLLFWPALFLDFPHDLRSFSSASNLNMLFSGIGASALCFATWN-YSVGV	243
QY	255	GAFFNLSMLTSDMAVAIRFIYQNVDMVLYLAFLVVGVIYTKTEKDPNTRALE	314
DB	244	GAVKSSAYIMYFVITIAASVLLIHENTWIALGGALTLLGLYI---SELKP-KAKLLE	299
QY	315	NG 316	
DB	300	NG 301	

RESULT 12
 COXI_TRYBB
 ID COXI_TRYBB STANDARD; PRT; 549 AA.
 AC P04371;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 GN CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
 GE COXI OR COI.
 OS Trypanosoma brucei brucei.
 OC Mitochondrion.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 RP [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 85037915.
 RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C., van Boom J.H., Benne R.;
 RT "The sequence of the gene for cytochrome c oxidase subunit I, a
 RT frameshift containing gene for cytochrome c oxidase subunit II and
 RT seven unassigned reading frames in Trypanosoma brucei mitochondrial
 RT maxi-circle DNA.";
 RL Nucleic Acids Res. 12:7327-7344(1984).
 RN [2]
 RP SEQUENCE OF 1-402 FROM N.A.
 RC STRAIN-164.
 RX MEDLINE; 85240413.
 RA Payne R.A., Rothwell V., Jasmer D.P., Feagin J.E., Stuart K.;
 RT "Identification of mitochondrial genes in Trypanosoma brucei and
 RT homology to cytochrome c oxidase II in two different reading
 RT frames.";
 RL Mol. Biochem. Parasitol. 15:159-170(1985).
 RN [2]
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2O) +


```
FT VARIANT 486 486 D -> N (IN GS).
FT FTID-VAR_007116.
FT VARIANT 496 496 G -> C (IN GS).
FT FTID-VAR_007117.
FT VARIANT 561 561 MISSING (IN GS).
FT FTID-VAR_007118.
FT VARIANT 588 588 A -> V (IN GS).
FT FTID-VAR_007119.
FT VARIANT 630 630 G -> V (IN GS).
FT FTID-VAR_007120.
FT VARIANT 655 655 R -> H (IN GS).
FT FTID-VAR_007121.
FT VARIANT 655 655 R -> L (IN GS).
FT FTID-VAR_007122.
FT VARIANT 728 728 A -> T (IN GS).
FT FTID-VAR_007123.
FT VARIANT 741 741 G -> R (IN GS).
FT FTID-VAR_007124.
FT VARIANT 850 850 L -> P (IN GS).
FT FTID-VAR_007125.
FT VARIANT 955 955 R -> Q (IN GS).
FT FTID-VAR_007126.
FT CONFLICT 459 460 AG -> VV (IN REF. 2).
FT CONFLICT 766 766 E -> D (IN REF. 2).
FT CONFLICT 807 807 V -> GARPSVSGAL (IN AAC50355).
SQ SEQUENCE 1021 AA; D/ECES3DA6233821 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 1021;
Best Local Similarity 21.3%; Pred. No. 1.9;
Matches 44; Conservative 33; Mismatches 78; Indels 52; Gaps 9;

QY 110 ILDCWTIV-----W-----SIITWFLGKTKYSVQVFGAACV-----GGLLLV 149
DB 147 MLNMGVILYLRPLWITQAQIGVLTWIIILSVTVTSITGLSISATNGKVGSGTFL 206
QY 150 LLSDSGVTAAGSNPLGDLFVITGSIILFTSTVG-----QEYCVRRKRIEYVAMI 200
DB 207 ISKSLGPELGGS---IGLIFAFANAGVAMHTVGFETVRLDQEQGAPVDPINDIRI 263
QY 201 GFEGMLISATEITVLERNALSSMOWSTG---LLAAVYVYALSFLCTLTPLLLKWSGNA 257
DB 264 GV-----VSTVLIALSLAGMEWESKAQVLFVLVIMVSFANVLVGTLPSPSEKASKG 316
QY 258 FPNLSMLTSDMMAVAIRTFIYNQEDV 284
DB 317 FFS---YRADI-----FVQNLVDPW 333

RESULT 14
YCBAY_CLOK
ID YCBAY_CLOK STANDARD; PRT; 311 AA.
AC P38943;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 33.5 KDA PROTEIN IN CAT1 5'REGION (ORFY).
OS Clostridium kluyveri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 555;
RX MEDLINE; 96146540.
RA Soehling B., Gottschalk G.;
RT "Molecular analysis of the anaerobic succinate degradation pathway in
CL J. Bacteriol. 178:871-880(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PCBM) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC -----

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L21902; AAA92345.1; -
CC PFAM; PF00892; DUF6; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 265 285 POTENTIAL.
SQ SEQUENCE 311 AA; 33507 MW; A5E7CBE9EC9ADF0B CRC64;

Query Match 6.0%; Score 103.5; DB 1; Length 311;
Best Local Similarity 19.1%; Pred. No. 0.76;
Matches 48; Conservative 51; Mismatches 121; Indels 31; Gaps 8;

QY 81 WYALGFDVHV-GNYLVNKAPELTSITSVSLDCWTVIWSIITWMLGKTKYSVQVFGA 139
DB 70 YLALGAILAVSISMSMLOFAVYTKASTAAVLFCTNAVFTIPFAYFILKEKINGITIVSI 129
QY 140 AICVGGLLLV-----LLSDSGVTAAGSNPLGDLFVITGSIILFTSTVGQECVRRKRI 194
DB 130 IVSLIGVVIIFNPVKMEGICG---GSRDLIGICFALVAAVVWSL-----YTVISKRI 179
QY 195 EVVA-----MIGVFGMLISATEITVLERNALSSMOWSTGLLAAVY-VYALSFLPC 244
DB 180 EIYGVGVFNCISFFFGVIALIL---LWVTGRPIFSGITLNNILVLLVLMGIFKAVGYIC 236
QY 245 TITPFLKMGGAFFNLSMLTSDMMAVAIRTFIYNQEDVLLYLAFLCVLVGVGFIYTKTE 304
DB 237 YLG--AIKETSATVASTVFLKPALATVLAIIIGSEIYVNVVIGVFIIGSIYSSN 294
QY 305 KDPNNTALEN 315
DB 295 KYANDLKVAN 305

RESULT 15
YMB8_YEAST
ID YMB8_YEAST STANDARD; PRT; 393 AA.
AC Q03730;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 43.7 KDA PROTEIN IN OST6-PSP2 INTERGENIC REGION.
GN YML018C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO YEAST YDR438W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z46659; CAA86637.1; .
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 346 366 POTENTIAL.
SQ SEQUENCE 393 AA; 43748 MW; DEC5C3EEF293F5ED CRC64;

Query Match          6.0%; Score 103.5; DB 1; Length 393;
Best Local Similarity 21.9%; Pred. No. 0.96;
Matches 63; Conservative 46; Mismatches 122; Indels 57; Gaps 13;

QY 63 LYGTILLYRRHKFLVPWYVYALLGFDVHGNYLVNKAFAELTSITSVSLDCWTIVWSIIF 122
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 123 LYETIKL--SAEFCILWF-----TANLVNLAFTSVASQTILS-----TTSFF 166

QY 123 TWMFLGT-----KSYVQFVGAICVGGLLLVLLSD-----SGVTAAGSNPLL 165
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 167 T-LFIGAICHVESLSKSVLGFSFVGIIMVTKSDSHQYQYQYQYQYQYQYQYQYQY 225

QY 166 GDFLVITGSIIF-TLST-----VGQYCVKRRDRIEVWAMIGVFGMLISATEITVLERNA 219
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 226 GNLLALAGAVLYGVYSTLLKREVGDETRVNMK---IFFGVGLFNLLFLWPSLIVLDFG 282

QY 220 LS--SMQWSTGLLAAYVYVALSSFL--PCTLTPFLKXGGAFFNLSMLTSDMNAVAIRT 275
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 283 WEPFSLPKDPKVVVIFVNCILITFVSDFCWAKAMLLTSLPTVTVGLSI-----TIPLAM 336

QY 276 F---YNOEVDMLYLAFLCVVYVVGIFIVTKTEKDPNNTRALENGNLD 319
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 337 FGDVIFKHKTMSALYLFGATLILGSPFFIINKSSEEHFENSITASNYE 384
```

Search completed: August 1, 2000, 21:32:27
Job time: 6047 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:47 ; Search time 51.03 Seconds
(without alignments)
406.259 Million cell updates/sec

Title: US-09-147-955-10
Perfect score: 1723
Sequence: 1 MSSSSRRWRENGMRTLL.....GNLDHEYSLLDQDDTPRKP 335

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_64:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428	24.8	505	T41303	hypothetical prote
2	145	8.4	324	S62543	hypothetical prote
3	134.5	7.8	292	D71181	hypothetical prote
4	125.5	7.3	370	S69718	hypothetical prote
5	124.5	7.2	305	F70030	conserved hypothet
6	124	7.2	330	H69798	conserved hypothet
7	122.5	7.1	341	T40424	hypothetical prote
8	121	7.0	306	J50286	membrane protein y
9	120.5	7.0	411	T03836	phosphate/phosphoe
10	119	6.9	490	A46391	CAMP receptor subt
11	114	6.6	311	A72487	hypothetical prote
12	114	6.6	567	S58750	NADH dehydrogenase
13	112.5	6.5	294	C72378	sugar ABC transpor
14	110.5	6.4	267	C75131	hypothetical prote
15	110.5	6.4	340	T01949	hypothetical prote
16	110.5	6.4	445	D69902	sodium-dependent t
17	109.5	6.4	541	B70769	hypothetical prote
18	109	6.3	442	S50332	NADH dehydrogenase
19	108.5	6.3	330	S76408	hypothetical prote
20	108.5	6.3	383	T38194	hypothetical prote
21	107	6.2	618	G75113	NADH dehydrogenase
22	106.5	6.2	531	S30011	probable membrane
23	106	6.2	390	T04096	glucose-6-phosphat
24	106	6.2	402	T14438	phosphate/triose-p
25	105.5	6.1	378	G70435	hypothetical prote
26	105.5	6.1	434	T40700	phosphate-phosphoe
27	105	6.1	306	C60749	conserved hypothet
28	105	6.1	549	ODUTWB	cytochrome-c oxida
29	105	6.1	617	F71019	hypothetical prote

ALIGNMENTS

RESULT 1

T41303
hypothetical protein SPCC320.08 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C:Accession: T41303

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Wambutt, R.; Wedler, E.
submitted to the EMBL Data Library, March 1998

A:Reference number: Z21986

A:Accession: T41303

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-505 <MO>

A:Cross-references: EMBL:AL022245; PIDN:CAA18310.1; GSPDB:GN00068; SPDB:SPCC320.08

A:Experimental source: strain 972h-; cosmid c320

C:Genetics:

A:Gene: SPDB:SPCC320.08

A:Map position: 3

C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC320.08

Query Match 24.8%; Score 428; DB 2; Length 505;
Best Local Similarity 31.4%; Pred. No. 3e-27;
Matches 93; Conservative 65; Mismatches 108; Indels 30; Gaps 7;

QY	24	LGQLVGFDAIMTF	SASLVSTVVDAPL	TMSFTYTVVALLY	GTILLYR-----R	72
DB	147	LGQVLSLCITATNT	FNCGYSGISNIPAF	QTFPLVALLTLVTP	TYTFVFMGFKYFEMIFR	206
QY	73	HKELVPWYW-YALL	GFVDVHGNYLVNKAF	ELTSVTSILDCWTIV	SIITWNFLGTKY	131
DB	207	HG-----WKYII	FAFDVEGNYFVYLA	YQYTNMLSLDSWAT	VAVILSFILKRV	260
QY	132	SVYQFVGAATCVG	LLLVLLSDSGVTAGS	-----NPLLDGFL	VITGSTLFTLSTVG	187
DB	261	HWSQLGVVACIGL	VLLVLSVD--VISR	GDYSANVPGLDGCM	IIGATCGVSNLTLEYF	318
QY	188	VKRDRIEVVAMIG	VGMUISATEITVLER	NALNSMOWST---G	LLAAYVYVYALSSFLP	244
DB	319	ASKLPLVYVIGQL	SLYSGISITSIQTIF	IFDRHHLTLRHTSE	MGGYLAGFI---LVN	375
QY	245	TLPLFLLKSGA	FFNLSMTDMVAIR	TFYVQEVDTWLYLA	FCVLVVVGIFTY	300
DB	376	SLAPILFRMSSA	TFYVNLSTSDTWSL	VIGIHVFYGHVWLY	PIAFVLIILGULFVY	431

RESULT 2

S62543

hypothetical protein SPAC12G12.12 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000

C:Accession: S62543; T37589

R:Devlin, K.; Odell, C.; Churcher, C.M.
Submitted to the EMBL Data Library, November 1995

A:Reference number: S62532

A:Accession: S62543

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-324 <DEV>

A:Cross-references: EMBL:Z66568; MID:g1052518; PID:CAA91507.1; PID:g1052530
R:Devlin, K.; Odell, C.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

Submitted to the EMBL Data Library, November 1995

A:Reference number: Z21727

A:Accession: T37589

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-324 <DR2>

A:Cross-references: EMBL:Z66568; PID:CAA91507.1; GSPDB:GNO0066; SPDB:SPAC12G12.12

A:Experimental source: strain 972h; cosmid c12G12

C:Genetics:

A:Gene: SPAC12G12.12

A:Map position: 1L

A:Inserts: 19/2; 254/2

Query Match 8.4%; Score 145; DB 2; Length 324;

Best Local Similarity 23.8%; Pred. No. 0.00014;

Matches 67; Conservative 42; Mismatches 117; Indels 56; Gaps 10;

QY 68 LLYRRHKFLVPWY-----WYALGFDVHGNVLYVKNKAFELTSVTSILDCWT 115

DB 53 VLYKRHS-QGRGYESLDHLPKHKVFNALPAINDICGSTMNVGLLYTSASTYQMTGRSL 111

QY 116 IWWSIIFTWMEFLGKYSVYQVGAACVGGLLLVLLSDSGVTAAGSNPLLGDFLVTGSI 175

DB 112 IIFVALFATLLKRTIGQLWLSFVVLGVAIVGVSGSS-SSIGSNPILGITAVLIGQF 170

QY 176 LFTLSTVGQECYVK--RKDRIVVAMTGVFGMLISATETVTLERNALSSNQWSTGLLAAY 233

DB 171 FLATOFTEEYILFSIQVDPSELVAYEGTYGVFFVLGMII-----SYYPFGSTTAGY 223

QY 234 -----VYVALSSFLFCTITPPLKMSGA-----AFFNLSML-----T 265

DB 224 HGMFDYSHVISRF--NEVPALYVIGSVILVSIAPFNVSGLATKLSATTSLLDIART 280

QY 266 SDNWAIVARTIYQVQWDLVLYLAFCLVVGVIYVYKTEKDP 307

DB 281 FGIWIIAMAGN--ESFHLQFLGLVILLIYGIFTYHSIIKFP 320

RESULT 3

D71181

hypothetical protein PH1729 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998

C:Accession: D71181

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekir

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137

A:Accession: D71181

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-292 <KAW>

A:Cross-references: GB:AP000007; MID:g3236134; PID:d1031786; PID:g3258160

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1729

Query Match

Best Local Similarity 22.0%; Score 134.5; DB 2; Length 292;

Matches 45; Conservative 51; Mismatches 80; Indels 29; Gaps 8;

QY 20 LGLGIGQLVSPDLAI---MTFSASLVSTTVDAPLTMSFTTYTVVALLYTILLYRRHKPL 76
DB 6 MNFALGVILAFSAFTWALTSVLKSVKMRVRSPLTNLFRLVIASGYLPLIYLN---L 62
QY 77 VP-----WY-----WYALGFDVHGNVLYVKNKAFELTSVTSILDCWTIWSIITWMP 127
DB 63 IFSKIMWVIVVPSGIIGP--MIGDMLFLEGKLLGVSRANMLTTLHLPIITWVLAHYLL 120
QY 128 GTKYSYVQVFGAAICVGGLLLVLLSDS---GVTAAGSNPLLGDFLVTGSIITLSTVQ 184
DB 121 GRPLNAYLFLAGLITLAILALLSEKSESGGINARGV-----LLVVAELLTIAIT 174
QY 185 EYCVKRDRIEVVAM---IGVFGML 206
DB 175 DMVSNPEPVLIALRISSGILGGL 199

RESULT 4

S69718

hypothetical protein YDR438w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 29-Oct-1999

C:Accession: S69718

R:Dietrich, F.S.

submitted to the EMBL Data Library, August 1995

A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461, 9831, and

A:Reference number: S69553

A:Accession: S69718

A:Molecule type: DNA

A:Residues: 1-370 <DI>

A:Cross-references: EMBL:U33007; MID:g927685; PID:YDR438w; GSPDB:GNO

C:Genetics:

A:Gene: YDR438w

A:Map position: 4R

C:Superfamily: probable membrane protein YML018c

Query Match

Best Local Similarity 7.3%; Score 125.5; DB 2; Length 370;

Matches 57; Conservative 57; Mismatches 102; Indels 45; Gaps 13;

QY 82 YALLGFDVHGNVLYVKNKAFELTSVTSILDCWTIWSIITWMPFLGK-----YSVQ 136

DB 124 FCVLMEV---ANLAANAALSYTTVASSTILSS---TSSFFT-LFLATSLGIETETSKKL 175

QY 137 VAAAICVGGLLLVLLSDSGV--TAAGSNPLLGDFLVTGSIITLSTVGQECVKKRD-R 193

DB 176 LGLFVSLFGIILIVMQSSKQODSVASFLVGNLTALLGSLGVSVVTTLLKYEISSKGLR 235

QY 194 IEVWAMIGVFGM-----LISATETVTLERNAL-SSQWSTGLLAAYVVALSFLFC 244

DB 236 LDIQMFGLYGVGIFTFLFPWFIILIDITHMETPELSPNFHISFLVNLNCIIIFVSDYFC 295

QY 245 ----TLTPPLKMSGAAPFNLSMTSDMNAVAITFTYNOEVDMLYLAFLVVGIFTY 300

DB 296 KALILTSPLVTVTALTFTPLAMPADFWREAFPT-----PW-YIIGVITFVSFFLV 347

QY 301 ----TKTEKDPNTRALENG 316

DB 348 NHRGESAVEKD---CAAVERK 365

RESULT 5

F70030

conserved hypothetical protein yvbY - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: F70030

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 350, 249-256, 1997

A:Authors: Foulger, D.; Eritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yanamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: F70030
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-305 <KUN>
A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PIDN:CAB15405.1; PID:ell86088
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvib

Query Match 7.2%; Score 124.5; DB 2; Length 305;
Best Local Similarity 23.3%; Pred. No. 0.0061;
Matches 74; Conservative 51; Mismatches 131; Indels 61; Gaps 13;

QY 27 LVSFDAIM--TFASLVSTVDAPLMTSTTTVVALLYGRLYYRHKFLVPMW-- 81

Db 12 LFAF-LVIMWGVNPLSKAALASPLPAGIRTLIGGLLVIVLPRHKLRLKETWPI 70

QY 82 ---YALLGFDVHG-----NYLVKAFELTSITSVILDCWTVMSIITFMFLGTYKS 132

Db 71 YLVSAALLNITLYGTLGTLNLYPLAGF-----SAIVFQVPLMGVSNLIGESNF 122

QY 133 VYQFVGAACVCGLLVLLSDSGVTAAGSNPLGSDPLVIT--GSTILFTLVSTVGQYCV 192

Db 123 VMKVIILGFGAVAVISAAGG-----GHTSVIGVLLAGSVAWALGTVYMKKTGSRVD 178

QY 193 RIEVVA--IGVFGMLISA--TEITVLERNALSMQWSTGLLAAYVTVVLSFELCTLT 247

Db 179 SIWVALQLTIGSVFLLIGFWTE-----SFAIGQWTAPE-----ITSLLF--IS 221

QY 248 PFLKSGAFAFNLSMLTSDMWAVAIRF-----LYNQVDNLYLAFCLVTVVG 296

Db 222 VFVIALGWLIVFTL-VGSEASKVASYTFLPLISIVASSIFLHEPLTSLLAGLLIYV 280

QY 297 IFIYTKERDPNNTAL 313

Db 281 SICLVNTKSKAKAAAI 297

RESULT 6

H69798

conserved hypothetical protein yetk - *Bacillus subtilis*C:Species: *Bacillus subtilis*

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: H69798

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertel
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Evington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yanamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033

A:Accession: H69798
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-330 <KUN>
A:Cross-references: GB:299107; GB:AL009126; NID:g2632866; PID:ell82700; PID:g2633034
A:Experimental source: strain 168
C:Genetics:
A:Gene: yetk
C:Superfamily: hypothetical protein ydek

Query Match 7.2%; Score 124; DB 1; Length 330;
Best Local Similarity 20.7%; Pred. No. 0.0073;
Matches 66; Conservative 48; Mismatches 139; Indels 66; Gaps 9;

QY 20 LGIGLQGVSEFLAMTFSA-----SLVSTTVDAPLMTSTTTTV--VALLVGTILLR 71

Db 58 LSSGLRFLRSVVLMLLFCIEKFPALTKDVFLLVQSFVGFVFCISLLIGV---- 112

QY 72 RHKFLVPMWYVALLGFDVHGNYLVNKAFLTSITSVILDCWTVMSIITFMFLGTY 131

Db 113 -----QYTTCTESGILTTTPMLIGILSFFLLREKI 143

QY 132 SVYQFVGAACVCGLLVLLSDSGVTAAGSNPLGSDPLVIT--GSTILFTLVSTVGQYCV 188

Db 144 EKTLIGILLVCGVMAINLFGAGSQDGTPhALFGNMLIANVIGENALFTMA---KLLS 200

QY 189 KRDRIVVAMIGVFGML-----ISATEITVLERNALSMQWSTGLLAAYVTVVLSFELFC 244

Db 201 PHISALAIATFVSLRGFLFPFALFEASDFVSPTVLDWS-----YVLYA--LFV 251

QY 245 TLTPPELLKMSG-----AAFFNLMLTSDMWAVARTIYNOEVDNLYLAFCLVTVVGIF 299

Db 252 TVLAFLYTKSGTVKVPAGVSGIFTSVLPVSAVILSGVILKEPFVHFVIGIACVIGIF 311

QY 300 YTKERDPNNTAL 318

Db 312 TVIKKKQPDAPAAEETL 330

RESULT 7

T40424

hypothetical protein SPBC405.03c - fission yeast (*Schizosaccharomyces pombe*)C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C:Accession: T40424

R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, March 1999

A:Reference number: Z21928

A:Accession: T40424

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-341 <SEE>

A:Cross-references: EMBL:AL035655; PIDN:CAB38602.1; GSPDB:GNC0067; SPDB:SPBC405.03c

A:Experimental source: strain 972h-; cosmid c405

C:Genetics:

A:Gene: SPDB:SPBC405.03c

A:Map position: 2

C:Superfamily: probable membrane protein YML018c

Query Match 7.1%; Score 122.5; DB 2; Length 341;
Best Local Similarity 19.0%; Pred. No. 0.0099;
Matches 64; Conservative 65; Mismatches 121; Indels 87; Gaps 13;

QY 27 LVSFDAIMTFASLVSTVDAPLMTSTTTVVALLYGRLYYRHKFLVPMW---- 81

Db 15 LLLFVFLMLISFLTSLDD--DNFFSPFLITYINTGTFVY-----LIPWYFSEKKT 67

QY 82 -----VALLGFDV--HGNLVNKAFLT 103

Db 68 RKRLMSELSMYVESVHDSFNLGTRPNPLGFRQTAYLSIGFCIIWFAANYFSNSSLGFT 127

```

QY   220 ---LSSQMSTGLLAAYVVVYALSFELCTLTTPFLKMSGAFFNLSMLTSDMVAIVRTF 276
      :::: ||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB   200 GKLTAPLSPSLGFLAVGLAIFGSIINAINMYLMIRNVSPALATSYAYVNYPVAVLLGTG 259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   277 IYNQEDWLYYLAFCLVVGVIPT 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   260 LGGETLSKIETLALGVIVFAVLVT 284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :


RESULT          9
T03836
phosphate/phosphoenolpyruvate translocator TABPPT10 precursor, plastid - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: T03836
R:Fischer, K.; Weber, A.; Fluegge, U.I.
submitted to the EMBL Data Library, August 1996
A:Description: The phosphate/phosphoenolpyruvate translocator from tobacco.
A:Reference number: Z15109
A:Accession: T03836
A>Status: preliminary; translated from GB/EMBL/DDBBJ
A:Molecule type: mRNA
A:Residues: 1-411 <FIS>
A:Cross-references: EMBL:U66402; NID:g1778144; PID:AAB40648.1; PID:g1778145
C:Genetics:
A:Gene: TABPPT10
C:Function:
A:Description: catalyzes the transport of phosphoenolpyruvate and phosphate across th

Query Match              7.0%; Score 120.5; DB 2; Length 411;
Best Local Similarity    21.8%; Pred. No. 0.018;
Matches                  62; Conservative 55; Mismatches 105; Indels 63; Gaps 13;

QY   49 PLTMS---FTTYTVVALYGILLVRHKE---LVPMYYALLGPVDVH--GNLYLNKA 99
      :::: ||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB   139 PVTVTUVQRFGSVILVMTLNLKYRKTKSGAQLV-----AILPLAVVHTLGNTFNWS 193
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   100 FELTSITSVISLODWTIWMISIFTMFGLTKYSVYQPVGAACVGGLLLVLDSGSVTAA 159
      : :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   194 LGRVAVSETHTIKAMEPFPSVLVSAMGLEGFPIITWNSSLVPVIGGVALASUTEASFNA 253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   160 GSNPLLGDPLVTIGSILTLS--HWGEYCVKRRDKRIEVVAMICVFGLMISATEITVLER 217
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   254 G-----FWSAMASNLTNSRNVLKKFMVKREDSLDNITFSI.....ITIM-- 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   218 NALSSMQNSTGLLAAYVVVYALSFELCTLPFLKMSGAFFNLSMLTSDMVAIVRTFI 277
      : ||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB   296 -----SFFLLAPYFAEG----VKFTPAYLEAGV---NVNQLYTRSLIAALCFHA 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   278 YNOEVDWL-----YTLAFP-----LVVVVGFIYTKTEKDPPN 309
      : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   341 Y-QQVSYMILQRVSPVTHSLGNCKRVVVIVTSLFEFTPVPSPIN 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :


RESULT         10
A46391
cAMP receptor subtype 3 - slime mold (Dictyostellium sp.)
C:Species: Dictyostellium sp.
C>Date: 18-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 29-Oct-1999
C:Accession: A46391
R:Johnson, R.L.; Saxe III, C.L.; Gollop, R.; Kimmel, A.R.; Devreotes, P.N.
Genes Dev. 7, 273-282, 1993
A>Title: Identification and targeted gene disruption of CAR3, a cAMP receptor subtype
A:Reference number: A46391; MUID:93170665
A:Accession: A46391
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <JOHG>
A:Cross-references: GB:S5235; NID:q265733; PID:N: AAB25437.1; PID:g265734
A>Note: sequence extracted from NCBI backbone (NCBIN:125367, NCBP:125369)
C:Keywords: cAMP binding; G protein-coupled receptor; phosphoprotein; transmembrane
```


QY 183 QGEYCVRKDRIEVVAMI-GVFG--MLISATETIVLERNALSSNQWSTGLLAAYVVTALS 239
| : : : | : : : |
Db 157 GWIAKTNNAPQTIIAFIRNPSGSLPPTLTTELSTPYSGIAGLIILAGOLIINYSAI 216
| : : : | : : : |
QY 240 SFLLCTLPFFLLKMSGAAFNLSMITS---DMMAVAIRTFLYNEDVWLTYLAFCLVAVYG 296
| : : : | : : : |
Db 217 SRIDLSLATITPAIPAVAIGIALTGVPTLM-----HLLGFIFITIG 260
| : : : | : : : |

Qy	297	IFYTK	302
		:	
Db	261	TLNLR	266

T01949
hypothetical protein F1104.14 - *Arabidopsis thaliana*

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01949

C:Accession: T01949

A;Description: The sequence of A. thaliana Fl104.
submitted to the EMBL Data Library, October 1998
A/MD accession: U-360947; A/MD accession: U-360948; A/

A;Reference number: Z14466

A;Status: translated from GB/EMBL/DDBJ

A:Residues: 1-340 <ABU>

C:Genetics:

A; Introns: 62/2; 101/2; 230/3; 281/2

Query Match	6.48;	Score 110.5;	DB 2;	Length 340;
-------------	-------	--------------	-------	-------------

Matches 74; Conservative 54; Mismatches 122; Indels 81; Gaps

QY 32 LAIMTFASLVSTVDAPLTMSTTYTVALLYGTILLYRRHKFLVP---WYWYALLGF 87

Db 14 LIVSNMIAGMVNALVKKVLDGGINHMVIATYRLGISTL-----FLLPVAYFWERASLMQY 68

QY 88 VDVHGNLVNKA FELTSITSV--SILDCWTIVWSIIFTW--MFLGTKYSVYQFVGAAI-C 142

D**b** 69 F-----YLLGLSYTSATLGS~~A~~FWAIMPSTLTFVMA~~L~~IFGFEKLSLTKIGYGVVLGTLISL 123

QY 143 VGGLLLVL----LSDSGVTAAGSNP-----LLGDFLVITGSILF-TLSTVGQEYCVK 189

Db 124 VGGLLLTMYQGIPLTNSPEQAANSNNHTGHENWIKGCFLLTGVLFSWMLIQAKINVK 183

QY 190 RKDRIEVVAMIGVFGML-----ISATEITVLERNALSSMQWSTGLLAAY 233

Db 184 YPCPYSTVILSVFGTLQCALSLIKTRHLEDWILRDELTII-----TVVIAGV 232

QY 234 VVYALSSFLFCTL-TPFLKMSG---AAFFNL SMLTSDMWAVAIRTFIYNQEV DWLYLA 289

D_b 233 VAQGM-----CTVGMSWCIKQGPVSSSFSPVLS---ATVFDLILHREIYLGSVIG 284

Qy 290 FCLVVGIFY-----TKTEKDPNNT 310

D**b** 285 SVVVIGLYIFLWSRSKQIVECKIMKLPNT 315

Search completed: August 1, 2000, 21:27:51

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:51 ; Search time 51.03 Seconds
(without alignments)
567.550 Million cell updates/sec

Title: 05-09-147-955-12
Perfect score: 2475
Sequence: 1 MVQPHVILTFPAQGHINP.....EGGSSHNKRAFDVAKGF 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_64:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1207	48.8	467	2	T02238	glucosyl transferase
2	1090.5	44.1	456	2	T71402	probable glucosyl
3	1049.5	42.4	458	2	C71420	hypothetical prote
4	809.5	32.7	479	2	F71419	probable indole-3-
5	766	30.9	475	2	F71419	probable indole-3-
6	762	30.8	460	2	T00639	hypothetical prote
7	761	30.7	484	2	D71419	probable indole-3-
8	750.5	30.3	471	2	A54739	indole-3-acetate b
9	727	29.4	456	2	T00506	indole-3-acetate b
10	712	28.8	438	2	T00511	indole-3-acetate b
11	645	26.1	453	2	T00511	indole-3-acetate b
12	553.5	22.4	455	2	T00584	indole-3-acetate b
13	540	21.8	453	2	T45603	glucosyltransferas
14	525	21.2	440	2	T00583	indole-3-acetate b
15	525	21.2	451	2	T45604	glucosyltransferas
16	522.5	21.1	452	2	T12981	hypothetical prote
17	500	20.2	462	2	T01732	UTP-glucose glucos
18	496	20.0	476	2	T03745	glucosyltransferas
19	491	19.8	476	2	T03747	glucosyltransferas
20	485.5	19.6	449	2	T45605	glucosyltransferas
21	477	19.3	466	2	T07404	probable glucosyl
22	471.5	19.1	490	2	T46162	glucosyltransferas
23	466.5	18.8	438	2	T45602	glucosyltransferas
24	444	17.9	447	2	T12978	hypothetical prote
25	443.5	17.9	488	2	T07786	UDP-glucose glucos
26	436.5	17.6	478	2	T08395	UTP-glucose glucos
27	432.5	17.5	487	2	S41951	UTP-glucose glucos
28	432	17.5	449	2	S41950	UTP-glucose glucos
29	431	17.4	346	2	S41954	UTP-glucose glucos

30	427	17.3	478	2	T05423	probable glucosyl
31	423.5	17.1	347	2	T06371	probable UDP-glucu
32	423	17.1	478	2	A71417	hypothetical prote
33	421.5	17.0	420	2	T08005	flavonol 3-O-glucu
34	419.5	16.9	472	2	S39507	glucuronosyl trans
35	417	16.8	471	2	S01052	flavonol 3-O-glucu
36	414.5	16.7	481	2	T01850	UTP-glucose glucos
37	409	16.5	385	2	T12980	hypothetical prote
38	409	16.5	471	2	S01037	flavonol 3-O-glucu
39	408.5	16.5	433	2	S51767	glycosyl transfera
40	407.5	16.5	452	2	G71416	probable glucosyl
41	404	16.3	471	1	S08325	flavonol 3-O-glucu
42	398	16.1	507	2	T46161	glucosyltransferas
43	393.5	15.9	394	2	S41952	UTP-glucose glucos
44	392.5	15.9	287	2	S41953	UTP-glucose glucos
45	391	15.8	452	2	T00981	flavonol 3-O-glucu

ALIGNMENTS

RESULT 1
T02238
glucosyl transferase, jasmonate-induced - common tobacco
C:Species: Nicotiana tabacum (Common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 24-Nov-1999
C:Accession: T02238
R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.
A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspensio
A:Reference number: Z14633
A:Accession: T02238
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-467 <KQ>
A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901
A:Experimental source: strain BY-2
C:Genetics:
A:Gene: JIGT
C:Superfamily: flavonol O3-glucosyltransferase

Query Match	48.8%	Score	1207	DB	2	Length	467
Best Local Similarity	50.1%	Pred. No.	1e-82				
Matches	234	Conservative	87	Mismatches	132	Indels	14
Gaps	7						
QY	5	HVLTTFPAQGHINPALQPAKLVKMGIEVTFSTSIYAQSRMDEKSIILNAPK--GLNFIP	62				
Db	9	HVLIATFPQGGHINPSLQFSKLLINLVKVTLSLSLAFNR-----IKNLPKIEGLTFAP	63				
QY	63	FSGDFEGEDHSDPV-FYMSQLKCGSETVKKILITCSENGQIPITCLLYSIFLPWAAEV	121				
Db	64	FSDDYDGNFKGSDDDYHLFNSAIKSHGSEFIANLIKAKNGYPTFRTVIITLMDWAGSV	123				
QY	122	ARVHVPISALLSQSPATIDYVFNFGHYEKAMAN-ESNDPNWSIOLPGLLETDLPLS	180				
Db	124	AKKLHPSILFHIQATVFDIYYRFTNFANFYKNDSDQO--IIEPLGPSLSSDDFPS	181				
QY	181	FLPYGAKSLRVALPPFKELIDTDAETTKILVNTFDELEPALNAIGKPYGIGLPL	240				
Db	182	FV--FDDVKSNDAWVESIKRQIEILNSEENPRILVNTFDALNLRVLRKNTVMVGIGL	239				
QY	241	IPSAFGLGNDPLDASGGDLFONSPNSVYVIFSGLMNPSISOMIEISK	300				
Db	240	IPSPFLDEKDRDNFFAADMIESENNYMWLDARANKSVIIAEGSAEISSQWMEISQ	299				
QY	301	GLIDIGRPFLMWIKENEKEEENKKGICIELEKIGKIVPWCQSLEVLKHPSLGCVFVSH	360				
Db	300	GLLKGRPFLMWIRETLNGEKPE-EKLTCKDELEKIGRIVKWCQSMEVLKHSYGCFLTH	358				
QY	361	CGWNSALESACQPVVAFPPQWTDQMTNAKQVEDVKSGVRVNRINDEGVVSEIKRCIE	420				
Db	359	CGWNSLESASGVPVACPIWNDQICNAKLIQDWKIGVRVNAKGGIIRDEFOKICIE	418				

QY 421 LYMDGGEKGBELRNKAKWKLAREAVKGGSSHNKLNKAFIDVAKG 467
 Db 419 IVMGDAEGBELRNKAKWKLAREAVKGGSSHNKLNKAYVNECLLG 465

RESULT 2
 C71402
 probable glucosyltransferase - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 A:Variety: columbia
 C:Accession: C71402
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegeger, T.; Pohl, T.M.; Terry, N.; G.; avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B. Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Chalwatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana
 A:Reference number: A71400; MUID:98121113
 A:Accession: C71402
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-456 <BEV>
 A:Cross-references: GB:297335; NID:92244747; PID:e327440; PID:g2244766
 C:Genetics:
 A:Map position: 4COP9-4G3845
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match 44.1%; Score 1090.5; DB 2; Length 456;
 Best Local Similarity 47.3%; Pred. No. 5.1e-74;
 Matches 221; Conservative 85; Mismatches 132; Indels 29; Gaps 9;

QY 3 QPHVLTTPAQGHINPALQAFKALNKGIEVTFSTSIYAQSRMDKSIILNAPKGLNFI 62
 Db 11 RPHVLTTPAQGHINPALQANLHGHATVYSTAVSAHRMGCEP---STKGLSFAW 67

QY 63 FSDGFDEGSDHSDPVPYKSLKCGSEVKKII---LTCSENGPITCLLYSIFLPWAA 119
 Db 68 FTDFGDLGKSFEDQKLYMSELKCGSNALRIILKANDATETETPTGVIIYSVLVPWVS 127

QY 120 EVAREVHIPALLWSOPATILDIYFNHFGYKAMANEKSNPNWSIQPLGLPLETRDLP 179
 Db 128 TVAREFHLPTLLWIEPATVLDIYFYFNYSYKHLFDVE-----PIKLPKPLITTTGDL 182

QY 180 SFLLPYGAKGSLRVALPFPKELDTLDAETTPKILVNTFDELEPALNAIEGYKFGIGP 239
 Db 193 SFLOPSKA---LPSALVTLREHEALTESNPKILVNTFSALEHDLUTSVKELKAPITGP 239

QY 240 LIPSAFLGGNDPLDASFGDLFONS-DYMEMLNKPNSVSVYISFGSLMN--PSISQME 296
 Db 240 LVSS-----SEKTDLFKSSDEDYTKWLDKSLERSVIYISLGHADOLPE-KHME 288

QY 297 EISKGLIDIGRPFLLWIKENKGEENKKGICIELEKIGKIVPMCSQLVLAHPSIGC 356
 Db 289 ALTHGVLTNRPLVIRKNPEKKKNFLRLGSDR-GLVGVNCSQTAVLAHCAVGC 347

QY 357 FVSHCGWNSALESACGVPVAPQPTQMTNAQVEDVWKSQVVRINEDGVVSEIEK 416
 Db 348 FVTHCGWNSLESACGVPVAPQPTQMTNAQVEDVWKSQVVRINEDGVVSEIEK 407

QY 417 RCIELVMDGGEKGBELRNKAKWKLAREAVKGGSSHNKLNKAFIDV 463
 Db 408 RCLEKVMGGEAEEMRENAKKKAMAVDAAAGGSPDLNLKGFVDE 454

RESULT 3
 C71420
 hypothetical protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)

A:Variety: columbia
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
 C:Accession: C71420
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegeger, T.; Pohl, T.M.; Terry, N.; G.; avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B. Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Chalwatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana
 A:Reference number: A71400; MUID:98121113
 A:Accession: C71420
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-458 <BEV>
 A:Cross-references: GB:297339; NID:92244901; PID:e326935; PID:g2244911
 C:Genetics:
 A:Map position: 4COP9-4G3845
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match 42.4%; Score 1049.5; DB 2; Length 458;
 Best Local Similarity 46.5%; Pred. NO. 6e-71;
 Matches 226; Conservative 78; Mismatches 115; Indels 67; Gaps 15;

QY 4 PHVLTTPAQGHINPALQAFKALNKGIEVTFSTSIYAQSRMDKSIILNAPKGLNFI 61
 Db 12 PHFLVTFPPAQGHINPSLELAKLAGTISGARVTPAASISAYNRR-MESTENVPETLIFA 70

QY 62 PDSGDFGEGF-----DHSKDPVF--YMSQLKCGSEVKKIILTCSENGPITCLLYSI 113
 Db 71 TYSQHDHDFKSSAYSQKQDQATGFMSEMRGRKTELTELIEDNRKONRPFCTCVYTI 130

QY 114 FLPMAAEVAREVHIPALLWSOPATILDIYFNHFGYKAMANEKSNPNWSIQPLGLPL 173
 Db 131 LLTWVAELA-----LPSIFYHYFNCGYDAISEMANTPSSSIKLPSLPL 174

QY 174 ETRDLPFLPYGAKGSLRVALPFPKELDTLDAETTPKILVNTFDELEPALNAI-RGY 232
 Db 175 TVRDLPFSIVSNVYAFEL---LPAPREQDSLKEEINPKILINTFOELEPAMSSVPDNF 231

QY 233 KYGIGLIPSAFLGGNDPLDASFGDLFONS-DYMEMLNKPNSVSVYISFGSLMNPSI 292
 Db 232 KIVPVGPLL-----TLRTD-FSSRGXYEMLDTRKADSVLYVSGFTLAVLSK 277

QY 293 SOMEISKGLIDIGRPFLLWIKENKGEENKKGICIELEKIGKIVP 341
 Db 278 KQLVELCKALIQSRPFLWITDKSYRNKDEGEKEED-----CISSEPEELDEIGNVVS 332

QY 342 WCSQLVLAHPSIGCFVSHCGWNSALESACGVPVAPQPTQMTNAQVEDVWKSQV 401
 Db 333 WCDQFRVLNHRISIGCFVTHCGWNSLESACGVPVAPQPTQMTNAQVEDVWKSQV 392

QY 402 V--RINEDG--VVESEIKRICELVMDGGEKGBELRNKAKWKLAREAVKGGSSHNK 457
 Db 393 VMEKKEEGVVDVSEIRKICEVME--DKAEEFPGNATRNKDLAAEAVREGGSSFNHL 450

QY 458 KAFIDD 463
 Db 451 KAFVDE 456

RESULT 4
 E71419
 probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 A:Variety: columbia
 C:Accession: E71419
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegeger, T.; Pohl, T.M.; Terry, N.; G.; avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgdomenech-erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Anstas, C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana
A:Reference number: A71400; MUID:98121113
A:Accession: F71419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-479 <BEV>
A:Cross-references: GB:297339; NID:g2244901; PID:e326931; PID:g2244906
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 32.7%; Score 809.5; DB 2; Length 479;
Best Local Similarity 38.4%; Pred. No. 5.9e-53;
Matches 187; Conservative 83; Mismatches 158; Indels 59; Gaps 15;

QY 5 HVLTTFPAQGHINPALQFAKNLVKMGIEVTFSTS-----IYAQSRMDEKSIILNAPKGL 58
DB 8 HVMLVSPGGGHVNPRLRLGLKLIASKGLLVTFVTEKPPWKKMQANKIQDGVLK-PVGL 66

QY 59 NFTP---FSDGDEGDFHSDKDPVFMYSQLRKCGSETVKIILTCSENGQIPITCLLYSIFL 115
DB 67 GFIRFEFFSDGFADDEKRFDFDAPRPHLEAVGKQEIKNLVK--RYNKEPVTCLINNAFY 124

QY 116 PMAEVAAREVHIPSPALLWSQATIIDIIYFNFGHYEKAMANESDNPWSIQLPGLPLET 175
DB 125 PWCDVAEELHIPSAVLWQSCACLTAYYYYHRLVAFPT--KTEPDISVEIPCLPLKH 182

QY 176 RDLPSL---LPVGKSGSRVALPPEKELIDTLDATETPK---TLVTFDELE----- 222
DB 183 DEIPSLPSPSTAFEGDI-----ILDQLRFENHKSFYIFDITFRELKIDMDHM 233

QY 223 ----PEALNAIEGYKYGIGLIPSAFLGNDPLDASGGDLFONSNDYMEWLNKSPNSS 278
DB 234 SQLCPQAI-----TSPVGPIFKNA-----QTLSSDVKGDISEPASCMEWLDSPSS 281

QY 279 VVYISGLMNPSSISOMEISKGLIDIGRPFLWVKENEKKEENKKGICIELEKIGK 338
DB 282 VVYISGFIANKQOEDELTAHGLSSGLSVLWVRPMEGTFFVPHVLP--RELKEGK 339

QY 339 IVPWCQSOLEYKHPSCFVSHCGWNSALESACGVPVVPAPQWTDQMTNAKQVEDVWKS 398
DB 340 IVEKCPQOERVLHAPALACELSHCGWNSTMEALTACVPVCPQWGDQVTDVAVLADVEKT 399

QY 399 GVRV-RINEDGVVESEI--KRCIELVMDGGKGEELKRNKAKKWLAREAVKEGSSHK 455
DB 400 GVRIGRAEMIVSREVVAEKLLEATV--GEKAVELRENARRWKEAEAAVADGSSDM 457

QY 456 NLKAFID 462
DB 458 NFKEFVD 464

RESULT 5
F71419
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Accession: F71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Gielavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgdomenech-erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Anstas, C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana
A:Reference number: A71400; MUID:98121113

A:Accession: F71419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <BEV>
A:Cross-references: GB:297339; NID:g2244901; PID:e326932; PID:g2244907
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 30.9%; Score 766; DB 2; Length 475;
Best Local Similarity 36.9%; Pred. No. 1.1e-49;
Matches 175; Conservative 91; Mismatches 172; Indels 36; Gaps 14;

QY 4 PHVLTTFPAQGHINPALQFAKNLVKMGIEVTFSTS-----IYAQSRMDEKSIILNAPKG 57
DB 8 PHVWLSVFFQCGHISPLRLGLKLIASKGLIVTFVTEPEPLCKMKRQANNQDGVLK-PVG 66

QY 58 LNFIPSPDGFDEGDFHSDKDPVFMYSQLRKCGSETVKIILTCSENGQIPITCLLYSIFLPW 117
DB 67 LGELRF-EFFEDGFVYKEDFDLLQKSLEYSVKREIKNLVKYK--QPVRCLINNAFVPW 123

QY 118 AAEVAREVHIPSPALLWSQATIIDIIYFNFGHYEKAMANESDNPWSIQLPGLPLET 176
DB 124 VCDIAEELQIPSAVLWQSCACLAAYYYYHOLVAFPT--ETEPEITVDVFPKPLTKHD 181

QY 177 DLPSFLTPGAGKSL-RVALPPFKELIDTLDATETPKILVNTFDELEPEALNAIE----G 231
DB 182 EIPSLPSPSSISIGTILEQIKRLHKPF-----SVLIETFOLEKDTIDHMSQLCPQ 235

QY 232 YKVGIGLIPSAFLGNDPLDASGGDLFONSNDYMEWLNKSPNSSVYISGLMNPSS 291
DB 236 VNFNPGLPTMA-----KTIKSDIKGIDSKPDSKIEWLDSPSSVYISGLTFLAK 290

QY 292 ISOMEISKGLIDIGRPFLWVKENEKKEENKKGICIELEKIGKIVPWCQSOLEYLKH 351
DB 291 QNOIDEIHAHGLSSGLSVLWVRPMEGTFFVPHVLP--ELEEKQIVWCQOERVLH 348

QY 352 PSLGCFVSHCGWNSALESACGVPVVPAPQWTDQMTNAKQVEDVWKSQVVR--INSDGV 409
DB 349 PAVACFLSHCGWNSTMEALTSGVPVICPQWGDQVTDVAVYMDVFKTGLRSLRGASDERI 408

QY 410 VESEI-KRCIELVMDGGKGEELKRNKAKKWLAREAVKEGSSHKNLKAFID 462
DB 409 VPREEVAERLLEATV--GEKAVELRENARRWKEAEASAVAYGGTSERNFOEFVD 460

RESULT 6
T00639
hypothetical protein F316.2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00639
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Arau
: Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z14197
A:Accession: T00639
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-460 <PED>
A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829862; GSPDB:GN00059; ATSP:F3
C:Genetics:
A:Gene: ATSP:F316.2
A:Map position: 1
A:Map position: 219/3
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 30.8%; Score 762; DB 2; Length 460;
Best Local Similarity 38.2%; Pred. No. 2e-49;
Matches 187; Conservative 78; Mismatches 163; Indels 62; Gaps 17;

QY 2 VQPHVILTPPAQGHINPALQAKNLVKMGIEVTFSTSIYAQSRMDEKSIILNAPKGLNFI 61
DB 8 VKGHVILPVPVQGHLPMPVQAFKRLVSNKVKVTTIATTTTASSITTPS-----LSVE 60
QY 62 PFSDFD-----EGFDRSKDPVYMSQLKCGSETVKKIIILTCSENGQIPITCLLYSTFL 115
DB 61 PISDFDFPIGIGF--SVDT--YSESEKLGSETLTLLIEKFSTDSPIDCLLYDSFL 116
QY 116 PMAEVAREVHIPSALLWSQPATILDIYFNHGYEKAMANESN-----DPNWS-IQL 167
DB 117 PWGLEVARSMELSAASFTNNLVJCSV-----LRKFSNGDFPLPADPNASPRRI 165
QY 168 PGLPILLETDRPLSPF-----LPYGAKSLRVALPPFKELIDTLDAAETTPKILVNTFDELEP 223
DB 166 RGLPSLSYDELPSFVGRHMLTHPEHG--RYLLNQFP-----NHENADMFLVNGFEGLE- 216
QY 224 EALNAIEG----YFYIGIPLSAPFALGGNDPLDASFGGDLFQON--SNDYMEWLSKPNSS 278
DB 217 ETQCENGESDAMKATILGPMIPSAIDORMEDDDKDYGASLKPISKCEMWELETKQAQS 276
QY 279 VYISFGLMNPISQMEIEISKGLIDIGRPFLLWIKENEKKEENKKGICIELEKIGK 338
DB 277 VAFVSFGSGLFELKQAEVAIALQESDLNFWLWIKAEHIAKLPE---GFVESTKDRAL 332
QY 339 IVPWCSOLEVLKHPSLGCFVSHCGWMSALESACGVPVYAFQWTDQMTNAKQVEDYWK 398
DB 333 LVSCNQLEVLHESIGCFTHCGWNSWMBESLSSGVPVYCCQWQDQVTDAYLIDVFKTCVRLGRGAT 414
QY 399 GYVRINEDG--VVESEIEKRCIELVMDGEGKEELRNKNAKKWELAREAVKEGGSHKN 456
DB 393 GYRAK-EEAGEVIVKSELVRLKGVME-GESSVKIRESSKKWKDLAVKAMSEGGSDRS 450
QY 457 LKAFIDDVAK 466
DB 451 INEFIESLCK 460

RESULT 7
D71419
C:Species: Arabidopsis thaliana (mouse-ear cross)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: D71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wamboldt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chaiwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113
C:Accession: D71419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-484 <BRV>
A:Cross-references: GB:297339; NID:g2244901; PID:e327472; PID:g2244905
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 30.7%; Score 761; DB 2; Length 484;
Best Local Similarity 36.9%; Pred. No. 2.6e-49;
Matches 178; Conservative 87; Mismatches 178; Indels 40; Gaps 15;

QY 5 HVILTPPAQGHINPALQAKNLVKMGIEVTF--STSIAQSRMDEKSIILNA-----PKGLNF 60
DB 13 HVMLVSFOGQGHVNPPLRLGLKASKGLLVTFVTELWKKRQANKVLDGLPVGSGS 72
QY 61 IPFSDGDFEGE-----DHSKDPVYMSQLKCGSETVKKIIILTCSENGQIPITCLLYSIPL 116

DB 73 TRP-EFFDEKAEODDRRDSFLYIAHLESVGIRESVSKLVRYREANEPVSCILNPPIP 131
QY 117 WAAEYAREVHIPSALLWSQPATILDIYFNHGYEKAMAN--ESNDPNMSIQLPGLPLE 174
DB 132 WYCHVAEENFIPCAVLAVQSCACFSAYH-----YQDGSVSFPTETEPDELDKVLCPCVPLK 187
QY 175 TRDLPSELPLGAKSLRVA-LPPFKELIDTLDAAETTPKILVNTFDELEPEALNAIEGY- 232
DB 188 NDEIPSFUHPSSRTGTRQAILGOFKNLSKF-----CVLIDSDSLEQEVVIDYMSSIC 241
QY 233 RYFGIGLIPSAFALGGNDPLDASFGGDLFQONSDYMEWLSKPNSSVYISFGSLMNP 292
DB 242 PVKTVGPLFKVA-----RTVTSDVSGDICKSTOKCLEWLSRPSKSVVYISFGTVAYLKQ 296
QY 293 SQMEIEISKGLIDIGRPFLLWIKENEKKEENKKGICIELEKI-----GKIVPWCSOLE 347
DB 297 EGIETIANGVLKSLGSLFWLWIRPPHDLKVETHVLP--QELKSSAKGKGMIVDMCPOEQ 354
QY 348 VLKHPSLGCFVSHCGWMSALESACGVPVYAFQWTDQMTNAKQVEDYWKSGVRV--RIN 405
DB 355 VLSHPSVACFVTHCGWNSWMBESLSSGVPVYCCQWQDQVTDAYLIDVFKTCVRLGRGAT 414
QY 406 EDGVVYESBEI-KRCIELVMDGEGKEELRNKNAKKWELAREAVKEGGSHKNKAFIDV 464
DB 415 EERVVPREEVAEKLEATV--GEKAEELRNALKWAEAAVAPGGSSDKNFRFEVKL 472
QY 465 ARG 467
DB 473 GAG 475

RESULT 8
A54739
Indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize
C:Species: Zea mays (maize)
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 29-Oct-1999
C:Accession: A54739
R:Szerszen, J.B.; Szczygłowski, K.; Bandurski, R.S.
Science 265, 1699-1701, 1994
A:Title: iaglu, a gene from Zea mays involved in conjugation of growth hormone indole
A:Reference number: A54739; MUID:94367368
A:Accession: A54739
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-471 <SSE>
A:Cross-references: GB:L34847; NID:g548194; PID:AAA59054.1; PID:g548195
C:Superfamily: flavonol O3-glucosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 30.3%; Score 750.5; DB 2; Length 471;
Best Local Similarity 35.5%; Pred. No. 1.5e-48;
Matches 172; Conservative 85; Mismatches 188; Indels 39; Gaps 13;

QY 4 PHVILTPPAQGHINPALQAKNLVKMGIEVTFSTSIYAQSRMD---EKSILNAPKGLNF 60
DB 3 PHVLVVPFGQGHNPVQPAKRLASKGVATLTVTRFTORTADVDAHVMVEA-----56
QY 61 IPFSDGDFE-GFDHSDKDPVYMSQLKCGSETVKKIIILTCSENGQIPITCLLYSIPLWAA 119
DB 57 --ISDGHDEGGFASAAAGVAEYLEKQAAAASLASILVEARASASDAFTCVVYDSYEDWVL 114
QY 120 EYAREVHIPSALLWSQPATILDIYFNHGY----EKAMANESNPNMNSIQLP----GLP 171
DB 115 PVARRMGLPAVPFSTQSCAVSAYYHFQSGRLAVPPGAAADGSDGAGAAALSEAFLGLP 174
QY 172 LLETDRDLSFLLPFGAKSLRVALPPFKELIDTLDAAETTPKILVNTFDELEPEALNAIEG 231
DB 175 EMERSELSFFVDFHGPPVPT--IAMQAIKQP---AHAGKDDWVLFNLSFEETEVLJAGLTK 229
QY 232 Y-KFYGIGLIP-----SAFLGGNDPLDASFGGDLFQONSDYMEWLSKPNSSVYISFGS 286
DB 230 YLKARAIGPCVPLPTAGRTAGAGRI--TYGANLVKPEDACTKWLDTKPDRSVAVVSFGS 287

```
QY 287 LMPSTISQMEIEISKGLIDIGRPFLLVWIKENEGKEENKKGICIELEKIG--KIVPWC 344
      : : ||::||: ||::||: : : : : : : : : : : : : : : : : : : : : :
Db 288 LASLGNAAQEEELARGLLAAGKPFLLVWV---RASDEHQVPRYLLAEATATGAAMVVP 343
      ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 345 QLEVLKHPSLGCFVSHCGWNSALESACGVVPAFFPQWTDQMTNAKQVEDVWKGVRVRI 404
      ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 344 QLDVLAHPAVGCFVTHCGWNSLLEALSGVPMVAMALWTDPTNARNVELANGAGVARR 403
      ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 405 NED-GVVESEETKRIELVMDGEGKEELRNKAKKWELEAREAVKEGGSHKMLKAFIDD 463
      ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 404 DAGAGVFLRGEYRCVRAMVDGGEAASAAKGAEMDRARAAPVAFGGSDRLDFVOF 463
      ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 464 VAKG 467
      : :
Db 464 VRAG 467
      : :

RESULT 9
T00506
Indole-3-acetate beta-glucosyltransferase homolog T20D16.11 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00506
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: 214159
A:Accession: T00506
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-456 <ROW>
A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642451
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A>Note: T20D16.11
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 29.4%; Score 727; DB 2; Length 456;
Best Local Similarity 37.7%; Pred. No. 8.3e-47;
Matches 178; Conservative 88; Mismatches 170; Indels 36; Gaps 17;

QY 3 QPHVLTTPAQGHINPALQFAK--NLVKGIEVTFSTSIYAQSRLMDEKSIINAPK--GLN 59
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 8 ETHVLMVTLPEFGHINPMLKLAHLSLSKKLHINLAT---IESARDLLSTVEKPRYPVD 64
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 60 FIFPSDFDEGHDKDPVFNYSQLRKCGSETVKKIILTCSENGQPTICLLYSIFLPAWA 119
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 65 LVFPSDGLPK--EDKAPETLTKSLUNKVGMNLKII-----EKKRYSCIIISPFPPVWP 117
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 120 EVAREVHIPSALLWSQPATILDIYFNFGHEKAMA--NESNDPNWSIQPLPLLETRDL 178
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 118 AVAASHNISCAILWQACGAYSVY---RYVMKTNFPDLEDLNTQVELPALPLEVRDL 174
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 179 PSLLPYGAKSLRVALPPFKELIDTDAETTPKILVNTFDELEPEALNAIEGK--FYGI 237
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 175 PSMLPSGGAHYNL-----MAEFADCL--RYVKWLVNYSFESEIESMADLKVPVPI 228
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 238 GPLIPSAFLGGNDPLDASFGDL--FONSND--YMEWLNKSPNSVYVYISFGSLMNPISQM 295
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 229 GPLV--SPFLLDGDEETLDGKNLDFCKSDCCMEWLDKQARSVYVYISFGSMLETLENQV 287
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 296 EETSKGLIDIGRPFLLVWIKENEGKEENKKGICIELEKIG--IVPWCQLEVLKHP 353
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 288 ETAKALKNRGLPFLVWIRPKA-----QNVAVIQEMKVGQGVVLEWSPQEKILSHEA 342
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 354 LGCFVSHCGWNSALESACGVVPAFFPQWTDQMTNAKQVEDVWKGVRVRI--DGWVES 412
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 343 ISCFVTHCGWNSMETVWAGVVPVAYPSWTQPIDARLLVDVFGVGRMENDSVDEGELKY 402
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 413 EETKRIELVMDGEGKEELRNKAKKWELEAREAVKEGGSHKMLKAFIDDV 464
```

```
Db 403 BEVERCIEAVTE--GPAAVDIIIRRAAEARKVARLALAPGSGSTRNLDLFISDI 453
      ||::||: ||: : : : : : : : : : : : : : : : : : : : : :
RESULT 10
T00507
Indole-3-acetate beta-glucosyltransferase homolog T20D16.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00507
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: 214159
A:Accession: T00507
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-438 <ROW>
A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642438
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A>Note: T20D16.12
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 28.8%; Score 712; DB 2; Length 438;
Best Local Similarity 38.4%; Pred. No. 1e-45;
Matches 177; Conservative 84; Mismatches 158; Indels 42; Gaps 18;

QY 14 QGHINPALQFAKNLVKMGIEVTFSTSIYAQ---SRMDEKSIINAPKGLNFIPESGDFE 69
      ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 7 QGHNLPMALFAKHLARTNLHFTLATEQAROLLSTADEP---HRPVDLAF--FSDGLPK 61
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 70 GFHSDKDPVFNYSQLRKCGSETVKKIILTCSENGQPTICLLYSIFLPAWAEEVAREVHIPS 129
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 62 --DPRDPDTLAKSLKKGAKNLKII-----EKKRFDCTIISVPFTFWPAVAAHNIPC 114
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 130 ALLWSQPATILDIYFNFGHEKAMA--NESNDPNWSIQPLPLETRDLPSFLLP--YGA 187
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 115 AILWQACGAFSVY---RYVMKTNFPDLEDLNTQVELPALPLEVRDLPSMLPSQGA 171
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 188 KGSRLVALPPFKELIDTDAETTPKILVNTFDELEPEALNAIEGK--FYGIGLIPSAFL 246
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 172 NVNTLMA-----EFADCL--KDVKWLVNYSFESEIESMSDLKPIIPGILV--SPFL 223
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 247 GGNDPLDASFGDLDFONSNDYMEWLNKSPNSVYVYISFGSLMNPISQMEIEISKGLD 306
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 224 LGNDEKTL---DMKAVDDYCMWLDKQARSVYVYISFGSLKSLNQVETIATALKNRG 280
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 307 RPFLLVWIKENEGKEENKKGICIELEKIG--IVPWCQLEVLKHPSLGCFVSHCGWN 364
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 281 VPFLWIRPKEG--ENVQV--LQEMVKEGKGVVTEWQGEKILSHMAISCFTHCGWN 335
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 365 SALESACGVVPAFFPQWTDQMTNAKQVEDVWKGVRVRI--DGWVESEIEKRCIELVM 423
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 336 STETVTVGVVPAVYPTWIDQPIDARLLVDVFGVGRMNDIAGELKVAEVCIEAVT 395
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
QY 424 DGEGKEELRNKAKKWELEAREAVKEGGSHKMLKAFIDDV 464
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:
Db 396 E-GPAAADNRRTATRELKHAARSAMSPGGSSAQNLDSFTSDI 435
      : ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||: ||::||:

RESULT 11
T00511
Indole-3-acetate beta-glucosyltransferase homolog T20D16.16 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00511
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: 214159
```

A:Accession: T00511
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-453 <ROU>
A:Cross-references: EMBL:AC002391; NID:g2642427; PID:g2642442
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 266/2
A:Note: T20D16.16
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 26.1%; Score 645; DB 2; Length 453;
Best Local Similarity 33.7%; Pred. No. 1.1e-40;
Matches 162; Conservative 93; Mismatches 160; Indels 66; Gaps 20;

QY 3 QPHVLTTPAQGHINPALQAFKLVKMGIEVTFSTSIYAQSRMDEKSLINAPKL-NFI 61
DB 8 ETHVLMVALPQGHINPMKFAKHLARTLNHFTLAT---IESARDLLSTDEPHSLVDLV 64
QY 62 PFDGCF--DEGFDHSDKDPVFMYSQLRKCGSETVKIILTCSENGQPIITCLLYSIFLPWAA 119
DB 65 PFDGLPRDPRDH--EPL--TESLRKVGANPFSKII-----EGKRFDCIISVPTPWVP 115
QY 120 EVAREVHIPSALLWSQATILDIYFNHGYEKAMA-NESNDPNKNSIQLPGLPLETRDL 178
DB 116 AVAAHNIPCAILMIEACAGFSYII--RYMYKTNFPDLEDPNQKVELPGLPFLVRDL 172
QY 179 PSFLIP-YGAGKSLRVALPPEKELIDTDAE-----TPPKILVNTFDEPEALNATEG 231
DB 173 PTLMPHGA-----IFNTLMAEVECLDKVKKWLANSFYELESVIESMFD 219
QY 232 YK-FYGGIPLPSAFICGNPDLPASFGG---DLFQNSDYMELWLNKSNSSVYISFGSL 287
DB 220 LKPIPIGPIVSPFLGADB--DKILDGKSLDMWKADDCYMEWLDQV-----SI 267
QY 288 MNPISQMEIEISKGLIDIGRPFVWIKENKKEENKKGICIEEL--EKIGKIIVPWCQ 345
DB 268 LKSSNQVETIATLAKNRGVPFLWVRPEKA-----ENVVLEDWVEEGQGVVLEMGQ 322
QY 346 LEVLKHPSLGCFVSHCGWNSALESGLAGVPVYAFPOWTDQNTNAKQVEDVWKSQVVRIN 405
DB 332 EKILCHMAISCFVSHCGWNSITETVSGVPVMAIYTFWQDILADLLVDVFGVGRMK-N 381
QY 406 E--DGWVESEIKRCIELVMDGGEELRNKNAKKWELAREAVKEGGSHKLNKAFID 463
DB 382 DVYDGLKVAEVEERICDAVTKGYD-AADMRRAALKQATRSAMAPGGSLARNLDLFIND 440
QY 464 V 464
DB 441 I 441

RESULT 12
T00584
Indole-3-acetate beta-glucosyltransferase homolog T27E13.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00584
R:Kounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.
A:Reference number: Z14177
A:Accession: T00584
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-455 <ROU>
A:Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150406
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 166/1

A:Note: T27E13.12
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 22.4%; Score 553.5; DB 2; Length 455;
Best Local Similarity 32.2%; Pred. No. 8e-34;
Matches 154; Conservative 78; Mismatches 182; Indels 65; Gaps 16;

QY 5 HVILTTPAQGHINPALQAFKLVKMGIEVTFSTSIYAQSRMDEKSLINAPK 56
DB 13 HVVAMPYFGRGHINPMNLCRLVRRYPNLHVTFTVEWLGFIGDPKDPRIHFSTLP- 71
QY 57 GLNFIPSDGFDGHDHSDKDPVFMYSQLRKCGSETVKIILTCSENGQPIITCLLYSIFLP 116
DB 72 --NLIP-----SELVRAKDFIGIDAVYTRLEPEPEKLL--DSLNSPPSVIFADTVI 121
QY 117 WAAEVAREVHIPSALLWSQATILDIYFN----FHGYEKAMANSNDPNKNSIQLPGLPL 172
DB 122 WAVRVGRKRNIPVYSLWMTSATILSFLHSDLLISHGHALFEPSEEEVVDY---VPLGSP 178
QY 173 LETRDLPSFLPYGAGKSLRVALPPEKELIDTDAETTPKILVNTFDEPEALNATEG- 231
DB 179 TKLRDLPLPIFDGYSR-VFKTAKLCFDEL-----PGARSLLFTTAYELEHKAIDAFTSK 231
QY 232 --YKYGIGIPLPSAFICGNPDLPASFGGDLFQNSN---DYMELWLNKSNSSVYISFGS 286
DB 232 LDPIYAIIGIPLIPEELS-----VONDKNKPNYIQWLEEQEGSVLYISQGS 278
QY 287 LNPISQMEIEISKGLIDIGRPFVWIKENKKEENKKGICIEEL-KIGKIIVPWCQ 345
DB 279 FLVSSEAQMEIEIVKLGRESGVRLWVARGE-----LKLKEAGSLGVVYVSWCDQ 329
QY 346 LEVLKHPSLGCFVSHCGWNSALESGLAGVPVYAFPOWTDQNTNAKQVEDVWKSQVVRIN 405
DB 330 LRVLCHKAVGFWTHCGFNSTLEGISGVPLAPFLFWQDILNKNKINVEDVWGMRIERT 389
QY 406 EDG--VWVESEIKRCIELVMD-GGEGEELRNKNAKKWELAREAVKEGGSHKLNKAFI 461
DB 390 KKNELLIGREIKVEVRFMDRESECKENRRRACDLSEISRGAVAKSSSNVNIDEFV 448
RESULT 13
T45603
glucosyltransferase-like protein - Arabidopsis thaliana
N:Alternate names: protein F12A12.180
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C:Accession: T45603
R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223008
A:Accession: T45603
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <CHO>
A:Cross-references: EMBL:AL133314
A:Experimental source: cultivar Columbia; BAC clone F12A12
C:Genetics:
A:Map position: 3
A:Introns: 161/1
A:Note: F12A12.180
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 21.8%; Score 540; DB 2; Length 453;
Best Local Similarity 31.4%; Pred. No. 8.1e-33;
Matches 153; Conservative 83; Mismatches 171; Indels 80; Gaps 19;
QY 6 VILTTPAQGHINPALQAFKLVKMGIEVT-FSTSIYAQSRMDEKSLINAPKGLNFIIFS 64
DB 10 VVLVPPAQGHISPMQWLAKTLHLKGFISITVQTKFNYSFSDDET-----HDFQFVITP 64
QY 65 DGFDFGFDHSDKDPVFMYSQLRK-CG---SETVKKIILTCSENGQPIITCLLYSIFLPWAAE 120


```

Db    160  PGLSPTRLSDLQ---ILUNGSHQVFNIFKSGELY-----KAKYLLFSAYELEPRAD 211
Qy    228  ATEG---YKFGIGPLIPSAFLG-GNDPLDASFGDDLQNSNDYMEWLNKSPNSVVVYIS 283
Db    212  FTSKFDPPVYSTGPIPLELSEVSGNE-----NRELDYFKWLDEQPESSVLYIS 260
Qy    284  FGSLANPSPISOMEIEISKGLDIDGRPLVMIVIKENSGKEENKKLGCIEELE-KIGKIVPW 342
Db    261  QGSFUSVSEAMEIIVGVREAGVFFWVARGGE-----LKKEALEGSLGVVYSW 311
Qy    343  CSQLEVLKHPHSGCFYSHCGWNSALESACGVPVVAFPQMTQMTNAKQVEDWVKSQVRY 402
Db    312  CDQLRVLCCHAAIGGFWMTHCGYSTNLEGICSGVPLLTTPVFWMDQFLNAKIVIEWRVGMGI 371
Qy    403  RINE--DGVVSEIEIKRICIELVMDG-GEKGELERKNAKKWKELAAREAVKEGSGSHKNLKA 459
Db    372  ERKKONELLYISDEIKELVRFMDGSEBEGEMRRRTCDLSEICRGNAVAKAGSSDANIDA 431
Qy    460  FIDDAK 466
Db    432  FIKDITK 438

RESULT 15
T45604
glucosyltransferase-like protein - Arabidopsis thaliana
N:Alternate names: protein F12A12.190
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
C:Accession: T45604

```

```

Accessed to the EMBL sequence database, December 1999
A:Reference number: 223008
A:Accession: T45604
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-451 <CHO>
A:Cross-references: EMBL:AL133314
A:Experimental source: cultivar Columbia; BAC clone FI2A12
C:Genetics:
A:Map position: 3
A:Introns: 160/1
A:Note: FI2A12.190
C:Superfamily: flavonol O3-glucosyltransferase

Query Match          21.2%; Score 525; DB 2; Length 451;
Best Local Similarity 30.8%; Pred No. 1..1e-31;
Matches 146; Conservative 34; Mismatches 198; Indels 56; Gaps

Qy 6 VILTTFPAGCHNPALQAFKNLYKMGIEVFTSTSYAQRNDEKSIILNAPKGLNFIPTSD 65
   ||| ||||| : || || | : || ||||| : || || | : || ||||| : || || | :
Db 10 VVLVAPQAQGHISIMOLAKTLHLKGFSTI-----AOTKFNIFSSDDTDFQCFVTFPE 64

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:19:53 ; Search time 236.66 Seconds
(without alignments)
2225.363 Million cell updates/sec

Title: US-09-147-955-11
Perfect score: 2105
Sequence: 1 agtgagcgcaagcaattaa.....acaaaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2105	100.0	2105	1 X02831	WO9905287 Seq ID 6
2	453.8	21.6	1506	1 X02826	WO9905287 Seq ID 1
3	439.8	20.9	1474	1 X02827	WO9905287 Seq ID 2
4	424.2	20.2	2062	1 X02828	WO9905287 Seq ID 3
5	289	13.7	1671	1 X02829	WO9905287 Seq ID 4
6	231	11.0	752	1 V31234	E. coli J96 pathog
7	231	11.0	7287	1 V02042	Plasmid pWRG3169 e
8	230.6	11.0	3699	1 V14340	Complete DNA seque
9	229.4	10.9	6046	1 V13845	Complete DNA seque
10	229.4	10.9	6245	1 V13844	Complete DNA seque
11	229.4	10.9	6448	1 V13843	Complete DNA seque
12	229.4	10.9	6580	1 V13167	Complete DNA seque
13	229.4	10.9	6613	1 V13842	Complete DNA seque
14	229.4	10.9	6824	1 Q39050	K.lactis/S. cerevi
15	229.4	10.9	6959	1 V13165	Complete DNA seque
16	229.4	10.9	6995	1 V13162	Complete DNA seque
17	229.4	10.9	7002	1 V13166	Complete DNA seque
18	227.8	10.8	3792	1 Q48463	Plasmid pg+host4 c
19	227.8	10.8	5234	1 Q48464	Plasmid pg+host5 c
20	227.8	10.8	6722	1 Q48465	Plasmid pg+host6 c
21	224	10.6	2885	1 T35417	Human amine transp
22	224	10.6	2885	1 V44908	Amine transporter
23	220.4	10.5	524	1 X30395	DNA encoding a hum
24	220	10.5	12494	1 V83206	Vector pHP-1 compr
25	219	10.4	608	1 X22239	Human secreted pro
26	218.4	10.4	616	1 X37426	Human secreted pro
27	216.2	10.3	6196	1 V13168	Complete DNA seque
28	216.2	10.3	6243	1 V13841	Complete DNA seque
29	216.2	10.3	6503	1 V13169	Complete DNA seque
30	216.2	10.3	7379	1 V13176	Complete DNA seque
31	216.2	10.3	8618	1 V18741	Complete DNA seque
32	216.2	10.3	8792	1 V18745	Complete DNA seque
33	215.2	10.2	1386	1 T30127	Hypoxanthine guani

C	34	213	10.1	8710	1 V32370	Complete sequence
C	35	211.4	10.0	1171	1 T84224	DNA encoding a Sta
C	36	211.4	10.0	5109	1 V13846	Complete DNA seque
C	37	211.2	10.0	5356	1 T43794	Plasmid PRIPHAT (r
C	38	210.8	10.0	4965	1 V18742	Complete DNA seque
C	39	208.8	9.9	7432	1 V61459	Vector pPR70 conta
C	40	207.6	9.9	814	1 V58695	3' fragment of pro
C	41	207.6	9.9	814	1 V61355	3' cDNA sequence o
C	42	196.8	9.3	4145	1 Q40279	Sequence of clone
C	43	196.8	9.3	6926	1 Q40281	Sequence of clone
C	44	193.8	9.2	2415	1 T13053	Cotton fibre-speci
C	45	193.8	9.2	2415	1 T30268	Cotton fibre clone

ALIGNMENTS

RESULT 1
X02831
ID X02831 standard; cDNA; 2105 BP.
AC X02831;
DT 14-MAY-1999 (first entry)
DE WO9905287 Seq ID 6.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Petunia hybrida.
FH Key Location/Qualifiers
FT CDS 341..1747
FT /*tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
PN WO9905287-A1.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 23-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI: 99-142940/12.
DR P-PSDB: W92952.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
PS Disclosure: Page 67-71; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.
CC Sequence 2105 BP; 569 A; 378 C; 429 G; 629 T;
SQ

Query Match 100.0%; Score 2105; DB 1; Length 2105;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	agtgagcgcaagcaattatgtgattagctcactcattagcagccagcgtttacac	60
DB	1	AGTGAGCGCAACGCAATTAATGTGAGTACCTACTTAGGACCCAGGCTTTACAC	60
QY	61	tttatgtctccgctcgtatgtgtgtggaattgtgagcggaataacaaatttcacacagga	120
DB	61	TTTATGTCTCCGCTCGTATGTTGTGGAATTGTGAGCGGATAACAATTTACACAGGA	120
QY	121	aacagctatgaccatattacccaagctcgaataaaccctactaaagggaacaaag	180
DB	121	AACAGCTATGACCTGATTAACCGGAGCTCGAAATTAACCTACTTAAGGGAACAAAG	180
QY	181	ctlgagctccacggtggtggtccgctctctagaaactagtgatcccccggtgcaggaat	240
DB	181	CTGGAGCTCCACGCGTGGCGGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAAT	240
QY	241	tcggtgtctgtcccaaatattacaacaaagaaataagcatccctttccccccttaa	300
DB	241	TCCGTTGCTGCTGCCCAATTTACAAACCAAGAAATTAAGCATCCCTTTTCCCCCCCTTAA	300

```
QY 301 aaacatcacagtttttaatttttccactaagcaagaaaaataaggcgagccctcatgtcoat 360
DB 301 AAAACATACAAGTTTTTAAATTTTTCACCTAAGCAAGAAAAATATGGTGCAGCCCTCATGTCTAT 360
QY 361 cttaacaacatttccagcacaaagcccatattaatccagcacttcatttgcacaagaactc 420
DB 361 CTTTACACACATTTCCACGACACAAAGCCCATATTATCCAGCAGCTTCATTTTGCACAGATCT 420
QY 421 tgcnaagatgggcataagagacatttttttccaaagcattttatgcccaaaagccgtatgga 480
DB 421 TGTCAAGATGGGCATAGAAGTGACATTTTCTACAAGCATTTTATGCCCAAGCCGCTATGGA 480
QY 481 tgaaaaatcccatcttaattgcacccaaaagattgaatttcattccatttccgatggctt 540
DB 481 TGA AAAATCCATCTTAAATGCACCAAAAGGATTTGAATTTTCATTTCCGATGGCTT 540
QY 541 tga tgaagggttttgatcattccaaagaccctgtattttacatgctcacaaacttcgtaaatg 600
DB 541 TGATGAAGGTTTGTGATCATTTCAAAGACCCGTGTATTTTACATGTCACAACTTCGTAAATG 600
QY 601 tggaaagtgaactgtcaaaaaataattctcacttgcctgaaatggacagccctataac 660
DB 601 TGGAAGTGA AACTGTCAAAAAAATAATTTCTACTTCTGCTGCTGAAATGGACAGCCCTATAAC 660
QY 661 ttgcctacttactccatttcttcccttggcgagcagaggttagcaagttccact 720
DB 661 TTGCCTACTTTTACTTCCATTTTCTCTTGGGCGAGCAGAGGTAGCACGTGAAGTTTCACAT 720
QY 721 ccccttgcctcttcttggagtcacacagcaaaatttgggacatatattacttcaactt 780
DB 721 CCCCTTCTGCTCTCTTTGGAGTCAACAGCAACAATATTTGGACATATATTACTTCAACTT 780
QY 781 tcatggatataaaaagctatgctaatgaatccaatgatccaaatttggtccattcaact 840
DB 781 TCATGGATATGAAAAGCTATGGCTAATGAATCCAAATGATCCAAATTTGGTCCATTTCAACT 840
QY 841 tccggggttccactactggaactcgagatcttccctcatcttttacttcccttatggctg 900
DB 841 TCCCGGGGCTCCACTACTGGA AACTCGAGATCTTCTCTTCAATTTTACTTTGCTTATGGTGC 900
QY 901 aaaaggaggtcttcgaattgcacttccaccattccaaagaattgatagacacattagatgc 960
DB 901 AAAAGGGAGTCTTCGAGTGCATCTCCACCTTTCAAGAAATTTGATAGACACATATAGATGC 960
QY 961 tgaacaccctctaagattcttggatatacatttggatgaattagagccctgaggcactcaa 1020
DB 961 TGA AACCCTCTTAAGATTCTTGTGAATACATTTGATGAATTAGAGCCTGAGGCCACTCAA 1020
QY 1021 tgaattgaaggttatagttttatggaattggaaccttgattcccttctgttcttctggg 1080
DB 1021 TGCAATGAAGGTTATAGGTTTATGGAATGAGCCGTTGATTCCTTCTGCTTCTTCTGGG 1080
QY 1081 tggaaatgaccctttagatcttcatttgggtggtgatcttttccaaaattccaaatgacta 1140
DB 1081 TGGAAATGACCCTTTAGATGCTTCATTTGTTGTTGATGCTTTTTCAAAATTCAAATGACTA 1140
QY 1141 tatggaatgggttaaacccaagccaaatccatcaagtgtgttatatatatttttggagctc 1200
DB 1141 TATGGAATGGGTTAAACCTCAAGGCCAAATTCATCAGTGTGTTTATATCTTTTGGGAGTCT 1200
QY 1201 aatgaatccatctattagccaaattggaggagatatacaaaaggtttgatagacataggag 1260
DB 1201 AATGAATCCATCTATTAGCCAAATGGAGGAGATATCAAAAGGGTTGATAGACATAGGAAG 1260
QY 1261 gccgtttttatgggtgataaaagaaatgaaaaggcaaaaggaagagaataaaaagct 1320
DB 1261 GCCGTTTTTATGGGTGATAAAGAAATGAAAAGGCAAAAGGAAGAGAGATAAAAAGCT 1320
QY 1321 tgggttgattgaagaatttgaaaaaataggaaaaatagttccattggtgttcacaacttga 1380
DB 1321 TGGTTGTATTGAAGAAATTGAAAAAATAGAAAAATAGTTCATGTTGCCATGTTTCCAACTTGA 1380
```

```
QY 1381 agttctaaaaacatccatttttaggatgttttttctctcatttggatggaaatccagcctt 1440
DB 1381 AGTTCTAAAAACATCCATCTTTTAGCATGTTTTTGTCTTCATTTTGGATGGAATTCAGCCTT 1440
QY 1441 agagagtttagcttggtagtgccagttgtggcatttccctcaatggacagatcaaatgac 1500
DB 1441 AGAGAGTTTACCTTGTGGAGTGCCAGTTTGTGGCATTTTCTCAATGGACAGATCAAAATGAC 1500
QY 1501 aaatgccaaaacaagttaagaatgtgtggaanaagtgagtgaaagtgagaataaaagaaga 1560
DB 1501 AAATGCCAAAACAAGTTGAAGATGTTGGAAAAGTGGAGTAAGAGTGAGAATAAATGAAGA 1560
QY 1561 tgggttcttgaagtggggaatccaaaagtgatttgaattggtaattggtagtgaggaga 1620
DB 1561 TGGTCTGTTTGAAGTGAGGAATCAAAAGGTGCTATTGAATTTGTAATGGTGAATGGAGGAGA 1620
QY 1621 gaaaggggaagaattggaagaagaatgctagaanaatgggaagaatggctagagaagctgt 1680
DB 1621 GAAAGGGGAAGAATTGAGAAAAGATGCTAAGAAATGGAAGAAATTTGGCTAGAGAAAGCTGT 1680
QY 1681 gaaggaagtggtatcttcacacaagaatttcaaaggtcttttattgtagatgttgccaaagg 1740
DB 1681 GAAGCAAGGTGATCTTCCACACAAGAATTTAAAGGCTTTTATTGATGATGTTTGCCAAAGG 1740
QY 1741 gttttaattattacaggcttttgcgtgatattacttcccttagttggcgattcactctt 1800
DB 1741 GTTTTAAATATTACAGGCTTTTGCCTGATATTACTTTCCCTAGTTGGCGATTCACCTCTT 1800
QY 1801 tctgacttgccttgacaaaaaactgagggaattgtctaaagcacacgcctaattgtcttaagaa 1860
DB 1801 TCTGCACTTGCTTGACAAAAAAGTGAAGGAATGTCTAAGACACGCTAATGCTTTTAAGAA 1860
QY 1861 gtcatttccaaaggttgaagcctgtctttaaacttatccaaacttatccagtaactcctaagggttc 1920
DB 1861 GTCATTTTCCAAAGCTTGAAGCTCTCTTTTAAACTTATTAGCCAGTAATCTATAGGGTTTC 1920
QY 1921 tctctattttctctgtctctctttttagccttttcttccaaaggtttaagaatagcg 1980
DB 1921 TCTTCTATTTTCTCTGTCTCTCTTTTATAGCTTTTCTTCCAAAGGTTTAAAGAATAGCG 1980
QY 1981 tgaacatagcttagtagctgtgtgtatctctctctctctctctaccagtgcaagattatgctt 2040
DB 1981 TGAACATAGCTTAGTACGTAGTCTGTGGTATCTCTATCTTACCAGGTGCAAGATTAATGCTT 2040
QY 2041 atgctgtcctcctcaaaattctttaaataaaatgcaagatgaaaagtgacaaaaaataaaaaa 2100
DB 2041 ATGCTGTCTCTCTAAATTTCTTAATAAAATGCAAGATGAAAAAGTACAAAAAATAAAAA 2100
QY 2101 aaaaa 2105
DB 2101 AAAAA 2105

RESULT 2
X02826 ID X02826 standard; cDNA; 1506 BP.
AC X02826;
DT 14-MAY-1999 (first entry)
DE W09905287 Seq ID 1.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Perilla frutescens.
FH Key Location/Qualifiers
FT CDS 17..1399
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
FT W09905287-Al.
FN 04-FEB-1999.
PD 16-JUL-1998; J03199.
PF 25-JUL-1997; JP-200571.
PA (SUNR ) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y, Yamazaki M;
```

DR WPI: 99-142940/12.
PT P-PSDB; W92947.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Disclosure; Page 49-52; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
CC
SQ Sequence 1506 BP; 364 A; 351 C; 449 G; 342 T;

Query Match 21.6%; Score 453.8; DB 1; Length 1506;
Best Local Similarity 61.6%; Pred. No. 4.9e-90;
Matches 870; Conservative 0; Mismatches 492; Indels 51; Gaps 7;

QY 337 aaattggtgagctcatgctatcattcattcaacaacattccagcaagggccattatattcc 396
DB 13 AAAATGTTCCCGCGCGCTGCTAGCAACGTTTCTTCCGCAAGGCCACATAAATCC 72
QY 397 agcaactcaattgccaagaatcttgcagatgggcatagaagtgcattttctacaag 456
DB 73 CGCCCTCCAAATTCGCCAAGAGACTCTTAAGAGCGGCACGTACATCTTTTCACGAG 132
QY 457 catttatgccaaagccgtatgataaataatcattcttaatgc-----accaaaagg 510
DB 133 CGTTTATGATCGCGCGCGATGGCCACACAGCTCCGCGCTCCGGAACCCACCGGG 192
QY 511 attgaatttcattccatttccgatgggtttgatgaagttttgatcattcaaaagccc 570
DB 193 CPTGACTTCGTGGCTTCTCCGAGCGGTACGACGACGGCTGAAGCCCTTGGCGGACGG 252
QY 571 tgcattttacatgcacaacttcgtaaatggaagtgaactgtcaaaataattct 630
DB 253 GAAGCGCTACATGTCGAGATGAAGCCCGCGCTCCGAGGCTTAAGAAACCTCTCTCT 312
QY 631 caattgctgtaaaatggacagcctaaacttgctactttactcatttccatttctcttg 690
DB 313 C-----AACACCACGAGGTACGCTTCGTCTACTTCCACCCTCTTTGCATG 360
QY 691 ggcagcagaggttagcaactgaagtgcacatcccttctgctctcttcttgagtcacaacgc 750
DB 361 GCGCGGAGGTGGCGCGGTGATGCTCCAGTCCCGAGCGCCCTTCTGTGGTGGAGCCGC 420
QY 751 acaaatattggacatatattctcaactttcatggatggaatgaagactatgctaatga 810
DB 421 CACCGTGTGTCATATATTACTTCTACTTCAACGCTACGACAGAGATCGACGCCG 480
QY 811 atccaatgatacaattggtccattcaacttccgggcttccactactggaactogaga 870
DB 481 TTCGACGAA-----ATTACGCTCCCTCGGCTTCCACCCCTGGAGCGCGAG 528
QY 871 tcttctcatttttacttctcttcttggtaaaaggagctcttcgagttgcacttccacc 930
DB 529 TCPTCCGACCTTCTGCTGCC-----GGAGACACCGGAGATTCGGTTGATGATG- --- 580
QY 931 attcaagaattgatagacacattagatgctgaaccacttccataagattcttgtaatac 990
DB 581 ----AAGGAGAAGCTGGAACCTTTAGACGGGTGAAGAGAAGCGGAAAGTGTGTGTAACAC 636
QY 991 atttgatgaattagagcctgagggcactcaatgcaattgaaggtttatgaattatgaat 1050
DB 637 GTTTGATGCTGGAGCCGATGCACTCACGGCTATTGATAGGTATGAGTTGATCGGAT 696
QY 1051 tggacattgattcctctcttcttcttgggtggaatgacccctttagatgcttcatctgg 1110
DB 697 CGGGCGTGTGATTCCTCCGCTTCTTGAGCGGGGAGATCCCTTCGGAACGCTTTACGG 756
QY 1111 tggatgatttttcaaaatcgaataaagcctta-----tggaaatgggttaaacctcaagcc 1164
DB 757 CGCGGATCTTTTGAAATAATCGAGGAGATAAATGCTGCTGGAGTGGTTGGACACGAGCC 816

QY 1165 aaattcatcagtggtttatatattcttttggagtcctaatgaatccattctattagccaat 1224
DB 817 GAAATCTTCGGTGGTGTATGTGTCGTTTGGGAGCGTTTGGAGTTTCAAAGACCAAT 876
QY 1225 ggaaggagatatacaaaaggttgatagacataggaagccgtttttatgggtgataaaga 1284
DB 877 GGAAGAGATTGGGAAGGCTATTAGCCTGCGGAAGGCCGTTTTTATGATGATACGAGA 936
QY 1285 aaatgaagaagcaaa---agaagaagaagaataaaaaagcttggttgtattgaagaatgga 1341
DB 937 ACAGAGAATACGACGCGCAAGAAAGAAAGAGAGTTGAGTTGCAATTGGGAATTGAA 996
QY 1342 aaaaaagaaaaatagttccatgggtgttcacaactgaagtttcaaaaaatccattcttt 1401
DB 997 AAAATGGGGAATAATGTTCTGCTGCTCGCAGTTGGAGGTTCTGCGCACCTCGCTT 1056
QY 1402 aggatgttttcttcttcttcttggatgaattcaagccttagagagtttagcttggagt 1461
DB 1057 GGGATTTTCGTGACGCATTTGGGTGGAACCTCGCTGTGGAGACTTGAGTTGCCGGGT 1116
QY 1462 gccagttgtgcttctcctcaatggacagatcaaatgacaaatgccaaacaaagttgaaga 1521
DB 1117 TCCGTTGTGGCGGTGCGCAGTGGTTTATCAGACGACGAATGCGAAGCTGATTGAGGA 1176
QY 1522 tgtgtggaagaagtggagtaagatgagaataaataagaatgggtgttcttgaagtgagga 1581
DB 1177 TGGTGGGGGACAGGGGTGAGAGTGAGAAATGAATCAAGGGGTGGGTTGATGGATCTGA 1236
QY 1582 aatcaaaaaggtgtattgaattggtgaatggatggagagagaagaaggaagaatgagaaa 1641
DB 1237 GATAGAGAGGTGTGTGGAGATGCTGATGATGGGGGTGAGAAGAGCAAACTGGTGAGAGA 1296
QY 1642 gaatgctaagaataagaagaatggtcctagagaagcgtggaagaagtggtgattctcaca 1701
DB 1297 AATGCCATTAATGAGACTTTGGCCAGAGAGCCATGGGAGAGATGGATGATCTTCACT 1356
QY 1702 caagaatttaaggcttttattgatgtgtgc 1734
DB 1357 CAAGAATCTCAAGCGCTTCTTCTCATCAAGTTGC 1389

RESULT 3

ID X02827 standard; cDNA; 1474 BP.
AC X02827;
DE WO905287 Seq ID 2.
KW Plant: flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Perilla frutescens.
FH key Location/Qualifiers
CDS 29..1360
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
FT WO905287-A1.
PN 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR) SUNTORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y,
PI Yamazaki M;
DR WPI; 99-142940/12.
DR P-PSDB; W92948.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
used to transform plants for improvement of plant coloration
PS Disclosure; Page 53-56; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have
flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
in the production of plant varieties with improved colour, for example in
the production of flower varieties with desired coloration and cut
flowers harvested from them.
CC
SQ Sequence 1474 BP; 369 A; 348 C; 438 G; 319 T;

5.

```
Qy 449 tctacaagcattttatgcccaaaagccgtagtgatgataaataatccattctttaaagcaccacaaa 508
Db 134 TTCCACGAGCGCTACGCATATGGCGCGCATG---TCCAGACCGCGCGCTGCTCAAAACGGG 190
Qy 509 ggaattgaatttcattccatttccgatggcttggatgaaggttttggatcattccaaagac 568
Db 191 CTCATCAATTTTGTGCTGCTTTCCGACGGGTATGACGAGGGTTACAGCCCGGAGACGAT 250
Qy 569 cctgtattttacatgtcaacaacttctgtaaatgtgggaagtgaacctgtcaaaaaataatt 628
Db 251 GGGAGAACTTACATTCGCGAGATGAAAGACGAGAGGTATAAAAGCCTTTGACGCGATACCTCT 310
Qy 629 -----ctcaacttgcctgaataatgacagcctataacttgctcacttacttactccatttcc 682
Db 311 GCAGCCAATAATGTCGATCAAAAGAACGACGAAATACAGTTGCGTGGTGTACTCCACCTC 370
Qy 683 ctctccttggcgagcagatgacgtgacgtgaagtccaatccctctctcttcttcttggagt 742
Db 371 TTTTCATATGGCGCGCCAAAGGTGGCGGTGAGTTCCTATCTCFCGAGCGCGCTACTCTGGATT 430
Qy 743 caaccagcaacaatatggacatatctacttcaacttcaacttctcatgatatgaataagctatg 802
Db 431 GAGCCAGCTAGGCTGTGATATATTTTACITTTTATTTCAACGGCTATACGCGAATC 490
Qy 803 gctaataatccaatgatccaaatggctccattccaactcccgggctcccaactctggaa 862
Db 491 GATCGGGTTTCGATCTATCTACATGCGC-----GGAGGACTCCCGAGTCTGGGCC 541
Qy 863 actcgagattctctcatttcttacttctcttatggcgcaaaaggagctctcgattgca 922
Db 542 CAGCGTATTTACCGTCTTCTCTCTCTT-----CCACGCGATGAGAGATTC 599
Qy 923 ctctccaccattcaagaatgtatagacacattagatgctgaacaccctcctaagattctt 982
Db 590 CGTCTACGTATGAGGAGAAATTTGGAACACTTTAGAAAGGTGAAGAAACCTAAGGCTTG 649
Qy 983 gtgaalacatttgatgaattagagctgagcactcaatgcaatgaagttaagtatt 1042
Db 650 GTGAACAGCTTTCATCGTTGGAGCTGATCGCTCAAGGCCATTGATAGTACGAGATG 709
Qy 1043 tatggaattgacccgttgattctctctcttcttcttggatgaaatgacctttagatgct 1102
Db 710 ATTGCAATCGGCGGCTGTGATCTCTCCGCACTTTCTTGGACGGTAAAGATCTCTCGGACAG 769
Qy 1103 tcatttgggtgattcttttcaaa-----aattcaaatgactatatggaatgatta 1153
Db 770 TCTTTCGCGGAGATTGTTTCGAGAAAGGTCGAATGACGAGCATTCGCTCGAATGGTTG 829
Qy 1154 aactcaagccaaattcatcagttgtttatatatcttcttggagctcaatgaatccatct 1213
Db 830 AGCACGAATCTTCGATCTTCGGTGGTTTACGTTTTCGTTGCGGAAGCTTCGTTAATACGACG 889
Qy 1214 attagcgaatgagagatatacaaaaggtttagacatatagaaagccgctttttatgg 1273
Db 890 AAGTCGCAATGGAAGAGATAGCAAGGGCTGTTAGATTCTGGAGGCGCGCTTTTGTGG 949
Qy 1274 gtgataaagaataatgaaaaagcgaagaagaagaataaaaaagcttggttgattgaa 1333
Db 950 GTGTGAAGATA-----AACGAGGAGGAAGAGGATTGTGAAGTTGCTGATGGAG 997
Qy 1334 gaattggaataataggaataatagttctcatgggtgttcacaacttgaagttctcaaaacat 1393
Db 998 GAGTTGAACAGCTGCGGAAATTTGTATCTTGTGTTCTCAATTTGGAAGTCTTGACGCAT 1057
Qy 1394 ccattctttagatgttttctctctctctctctctctcaatggacagatcaaatgacaaatg 1453
Db 1058 CCTCTGTGGGATGTTTCTGTGACACACTCGCGGTGGAAATCGATCTAGAGAGTATATCT 1117
Qy 1454 tgtggagtgccagttgtggcatttctcctcaatggacagatcaaatgacaaatgcaaaacaa 1513
Db 1118 TTCGGGGTCCGATGTGGCTTTTCCGAGTGGTTCGATCAAGGGACGAATGCGAGCTG 1177
```

```
Qy 1514 gttgaagatgtgtggaataagtgagtaagagtgagaataaataagatggt---gttgtt 1570
Db 1178 ATGAGGATGTGTGGAGGACGGGTGTGACAGTGACAGCTAATGAGGAGGAGGTAGCGTGGT 1237
Qy 1571 gaaagtgaagaaatcaataaagtgattgaattggttaattggatgagagagaaaggggaa 1630
Db 1238 GATGGTGATGAAATTAGGAGATGTTATGAGGAGGTTATGATGGGGGAGAAAAGAGTAGG 1297
Qy 1631 gaattgagaagaatcctaagaataatggaagaatggcttagaagctgtgaaggaaggt 1690
Db 1298 AAACCTTAGAGAGAGTCTGGCAAGTGGAGGATTTGGCAAGAAAGCTATGAGGAAGAT 1357
Qy 1691 ggaattccacacaagaatttaaggcctttatttattgatgtgtg 1733
Db 1358 GGATCTTCAGTTAAACAACCTCAAGGTCTTTCTTGATGAGGTG 1400

RESULT 5
X02829 standard; cDNA; 1671 BP.
AC X02829;
DE W0905287 Seq ID 4.
KW Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour; ds.
OS Torenia hybrida. Location/Qualifiers
FH Key 45.1481
FT /tag= a
FT /product= "protein with flavonoid 5-transglycosylation activity"
PI W0905287-A1.
PD 04-FEB-1999.
PF 16-JUL-1998; J03199.
PR 25-JUL-1997; JP-200571.
PA (SUNR ) SUNFORY LTD.
PI Gong Z, Kusumi T, Mizutani M, Saito K, Tanaka Y, Yamazaki M;
PI WPI; 99-142940/12.
DR P-PSDB; W92950.
PT Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be used to transform plants for improvement of plant coloration
PS Disclosure; Page 60-64; 89pp; Japanese.
CC This invention describes the isolation of plant proteins which have flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used in the production of plant varieties with improved colour, for example in the production of flower varieties with desired coloration and cut flowers harvested from them.
CC Sequence 1671 BP; 530 A; 344 C; 399 G; 396 T;
SQ
```

Query Match 13.7%; Score 289; DB 1; Length 1671;
Best Local Similarity 56.1%; Pred. No. 3.4e-54;
Matches 847; Conservative 2; Mismatches 557; Indels 105; Gaps 12;

```
Qy 329 aagcaagaataatggtgcagcctcatcttaacaacatttccagcaagggcat 388
Db 36 AAAAAAATGTTAACAAACCCCATATTCTACTAGCAACATTCCCAAGCAAGGCCAC 95
Qy 389 attaatccagcactccaatttggcagaagaatcttgcgaagtgggcatag-----aagtg 442
Db 96 ATAAACCTTCTCTCGAGTTTCGCCAAAAGGCTCCTCAACACGGGATAGTCGACCAAGTC 155
Qy 443 acatttctcaagcatttatgccaaagcctgatgataaataatccattcttaagca 502
Db 156 ACATTCTTCAGGAGTGTATACGATTTGACGCGCATCGCTTCGAAACCGCATCCGAGC--- 212
Qy 503 ccaaaagattgaatttcattccatttccgatggttggatgaaggttttgatcatca 562
Db 213 ---AGCAGAAATCGATTTTCGTGGCATKTYCAGATTCTTACGATGATGGCTTTAAGAAAGGC 269
Qy 563 aaagaccctgtattttacatgtcacaaactcgttaaatggaagtgaactgtcaaaaa 622
Db 270 GACGTGGCAAAACTACATGTCGGAGATGAGAAAGCGCGGCAACGAAGCCCTTAAAGAC 329
```

Qy	623	ataattctcaacttgctc-----tgaaaatggacagccctataacttgc	664
Db	330	ACTCTATTAAAGTCTCAACGATGCTCGATGGGAAGTGAATGTTACAAATCCGCTGAGCTTT	389
Qy	665	ctactttactccatttctctcttgggcagcagaggtagcacgtgaagttcacatccct	724
Db	390	GTGGTGTACTTCTCATCTATTTTCGTGGCAGCTGAAGTGGCGGTGAAGTCGACGTGCCG	449
Qy	725	tcgtctctctcttgagttcaacacgagcaacaatttggacatatatttacttcaactttcat	784
Db	450	AGTCCCTCTCTTGGAAATGAACCGGTACGGTTTTCGATGTGTACTATTTTACTTCAAAT	509
Qy	785	ggatatgaaaaagctatggtctaataatccaatgatccaaatggtccattcaacttccc	844
Db	510	GGGTATGCCGATGATATCGATCGGGCTCAGATCA-----AATCCAAACTGCCC	557
Qy	845	ggcttccactactggaaactcgagctctctctcaatttcaactcttattggtgcaaaa	904
Db	558	AATCTCCGAGCTCTCCAAAGCAAGATCTCCCTCTTTCCTACTCCCTT-----	606
Qy	905	ggaggtcttgagttgcacttccacatttcaaaaagaattgatagacacattagatgctgaa	964
Db	607	-CGAGCCCGCGAGATTCGGAACCCCTAATGAAGAAAAAGTTTCGACACGCTCGACAAAGAA	665
Qy	965	accactcttaagattcttctgatacatttggatgaattagagcctgaggcaactcaa	1024
Db	666	CGAAAGCGAAGGTCTTGATAAACAGTTCGACGCAATTAGAAACCGAACACTCAAGACC	725
Qy	1025	attgaagttataagttttatggaattggacgcttgatctctcttctgtct--tcttgggt	1081
Db	726	ATCGACAGGTATGAACATAATATCCATCGGCCCATTAATCCCATCATCGATATCTCAGAT	785
Qy	1082	ggaaatgaccttttag-----atgcttctattgggtggtgatctttttcaaaattcaaat	1135
Db	786	GGCAACGACCCCTCATCAAGCAACAAATCCTACGGTGGAGACCTCTTCAGAAAAAGCCGAT	845
Qy	1136	ga---ctatatgaatggttaactcaaaagcaaatctacagttggtttatatctttt	1192
Db	846	GAAACTTACATGACCTGGCTTAACCTGCAAAACCCGAATCTCGGTGCTGTTTACGTTTCGTTC	905
Qy	1193	ggagtgctaaatgaatccactctattatgccaaaatggaggagatatcaaaagggttgtagac	1252
Db	906	GGGAGCTCTCTGAGGCTCCGAAACCCCAAAATGGAAGAAATAGCAATAGGGCTTTTTCAGAC	965
Qy	1253	ataggaaagccgtttttatgggtgatacaaaagaaaatgaaaagcgaagagaagagaat	1312
Db	966	ACCAAAATGCCAGTTCTCTGGGTGATAGAGAAACAGGAAGGGCGGACAAACAGAGCAA	1025
Qy	1313	aaa-----aagcttggttgatatgagaatctggaaaaaataggaa	1351
Db	1026	GCAGAAGAAGAAGAGAGCTGCTGAGCTTCTTTGATCTCGAGAACTGAACGACTCGGG	1085
Qy	1352	aaaaatgttccatgggtttccacaactctgaagttctcaaacatctcttaggatgtttt	1411
Db	1086	AAATCGTGACATGGTGCTCACAATTGGATGTTCTGACGCATNAGTCGGTGGGATGCTTC	1145
Qy	1412	gtttctcatgttggatggaaattcagcctcttagagagtttagotttggagatgcccagttgtg	1471
Db	1146	GTGACGCATTTCGGGTTTGGAAATTCTGCTATCGAGAGCCCTGGCTTGTGTGTGTCGCCGTGGTG	1205
Qy	1472	gcatttctctcaatggacagatcaaaatgacaaatgccaaacaagttgaaagatgtgtggaaa	1531
Db	1206	TGCTTTCTCAATGGTTCGATCAAGSGACTAATGCGAAGATGATCGAAGATGTGTGGAGG	1265
Qy	1532	agtggagtaagctgagaataat---gaagatggtgttcttgaagtggagaaatcaaa	1588
Db	1266	AGTGGTGTGAGGATCAGATGTAATGAGGAGCGCGCTTGTGTATAGCGCTGAGATTAAAG	1325
Qy	1589	aggtgtaattgaaattggtcaattgdatggagggagagaaaggggaagaattcgagaagaatcgct	1648
Db	1326	AGGTGCCTCTCGGAGGTTTAAAGATCTCAGAG-----TTGAGAAAGAGCGCA	1373

Qy	1549	aagaaatcgaaagaattgctgctgagagaagctgtgaagggaag---gtgggtctcttcacacaag	1705
Db	1374	ATGATGTGGGAAGGGTTGGCTCAAGAAGACTATGGATGAAGAACGTGGGATCATCAATGAAC	1433
Qy	1706	aattaaagcctttattgatgatgtgctgccaaggggttttaatatattacagagcttttgcc	1765
Db	1434	AATCTGAAGAATTTTATTACTAGGATTATTAAATGAAATGCCTCATAGTTGTACTATAT	1493
Qy	1766	gtgatattact	1776
Db	1494	ATGTATTATT	1504
RESULT	6		
ID	V31294/C		
AC	V31294	standard; DNA; 752 BP.	
AC	V31294		
DT	01-OCT-1998	(first entry)	
DE	E. coli J96	pathogenicity island contig #108.	
KW	PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;		
KW	PAI V; pheV; vaccine; protective immune response; ds.		
OS	Escherichia coli.		
PN	W09822575-A2.		
PD	28-MAY-1998.		
PP	21-NOV-1997; U21347.		
PR	14-OCT-1997; US-061953.		
PR	22-NOV-1996; US-031626.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(UTWI-) UNIV WISCONSIN.		
PI	Choi GH, Dillon PJ, Welch RA;		
DR	WPI; 98-312461/27.		
PT	New isolated uropathogenic E. coli nucleotide sequences - used to		
PT	develop products for the detection of pathogenic E. coli and to		
PT	elicit an immune response to pathogenic E. coli		
PS	Claim 21; Page 221-222; 250pp; English.		
CC	This sequence represents a E. coli strain J96	contig containing	
CC	pathogenicity island (PAI) sequences, and represents a nucleic acid		
CC	molecule of the invention. PAIs are large fragments of DNA which comprise		
CC	pathogenicity determinants. The sequences of the invention are taken from		
CC	PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)		
CC	on the E. coli chromosome and is greater than 170 kb. PAI V is located at		
CC	approximately 94 min (at pheR) on the E. coli chromosome and is		
CC	approximately 160 kb in size. Antibodies specific to the proteins encoded		
CC	by the PAI open reading frames of the invention can be used in kits to		
CC	detect uropathogenic E. coli. The proteins are used in vaccines to elicit		
CC	a protective immune response in an animal to the uropathogenic E. coli		
CC	strain J96.		
QQ	Sequence	752 BP; 162 A; 213 C; 203 G; 172 T;	

Query Match 11.08; Score 231; DB 1; Length 752;
Best Local Similarity 99.60; Pred. No. 1.1e-41;
Matches 242; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy	1	agtgagcgcaacgcaaatatgtgagtagtcactcatttagcaccacccaggctttacac	60
Db	289	AGTAGCGGCACGCAAAATATGTGAGTTAGCTCATTAGCGACCCAGGCTTTACAC	230
Qy	61	tttatgcttcocgctcgatgtgtgtggaattgtgagcggaatacaatttcacacagga	120
Db	229	TTTATGCTTCCGGCTCGTATGTTGTGTGAAATGTGAGCGGATACAAATTTACACAGGA	170
Qy	121	aacagcatgacacatgatacgcacgaatcgaaatcaaccctcactaaagaggaacaaag	180
Db	169	AACAGCATGAGCCATGATTAGCGCAAGCTCGAAATTAACCCCTCATTAAAGGGGAACAAAG	110
Qy	181	ctggagctcca-cgcggtggcggcgcctctagaactagtggaatcccccgggctgcaggaa	239
Db	109	CTGGAGCTCCACCGGGTGGGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA	50
Qy	240	ttc 242	
Db	49	ttc 47	


```

RESULT 7
V02042 ID V02042 standard; DNA; 7287 BP.
AC V02042;
DE 08-JUN-1998 (first entry)
KW Interleukin-12; IL-12; cytokine; growth factor; mouse;
KW plasmid pWRG3169; cancer; tumour; metastasis; gene therapy; ds;
KW cyclic; circular.
OS Chimeric - Mus musculus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
OS Chimeric - Rhesus macaque polyoma virus.
FH Key Location/Qualifiers
FT promoter 1..628
FT /tag= a
FT /note= "CMV promoter"
FT IDNA 629..810
FT /tag= b
FT CDS 953..1673
FT /tag= c
FT /product= p35 subunit
FT /note= "contains an intron"
FT intron 1259..1331
FT /tag= d
FT polyA_site 1797..2024
FT /tag= e
FT /note= "bovine growth hormone polyA site"
FT promoter 2110..2737
FT /tag= f
FT IDNA 2738..2919
FT /note= "CMV promoter"
FT CDS 2983..3990
FT /tag= h
FT /product= p40 subunit
FT polyA_site 4075..4306
FT /tag= i
FT /note= "bovine growth hormone polyA site"
W09746263-Al.
11-DEC-1997.
PF 04-JUN-1997; U09591.
PR 05-JUN-1996; US-659206.
PA (AURA-) AURAGEN INC.
PI Rakhmillevich AL, Yang N;
WPI: 98-041898/04.
DR P-PSDB; W44004-05.
PT Interleukin-12 gene therapy of tumours - comprises delivering
PT construct containing promoter and sequences encoding interleukin-12
PT p35 and p40 subunit(s) to target cells in vivo
PT Claim 5; Page 24-30; 50pp; English.
PS Plasmid pWRG3169 is a tandem plasmid encoding both subunits, i.e.
CC p35 (see W44004) and p40 (see W44005), of murine interleukin-12
CC (IL-12). Each subunit gene was cloned from a mouse spleen cDNA
CC library and has been placed under the transcriptional control of
CC a separate cytomegalovirus (CMV) promoter. An SV40 splicing
CC donor/splicing acceptor is provided between each subunit gene and
CC its CMV promoter. The backbone of the plasmid is from pUC19.
CC pWRG3139 induces at least twice the expression of IL-12 as the
CC bicistronic vector pWRG3196 (see V02043) in vivo and in vitro. A
CC novel method of treating tumours in a mammal involves delivering
CC copies of an expressible foreign genetic construct, especially
CC pWRG3169 or pWRG3196, comprising a promoter operative in the
CC mammalian epidermal cells and DNA sequences encoding p35 and p40
CC subunits of IL-12 to target cells in vivo. Delivery of the
CC construct allows IL-12 expression for treatment of solid,
CC metastatic or disseminated tumours, and regression of established
CC tumours. The treatment is effective even when the genetic
CC construct is delivered to a site distant from the tumour.
CC Sequence 7287 BP; 1798 A; 1877 C; 1846 G; 1766 T;
SQ
Query Match 11.0%; Score 230.6; DB 1; Length 3699;
Best Local Similarity 99.6%; Pred. No. 2e-41;
Matches 242; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtatgctcactcattagcaccacccaggcttaacac 60
|||||
DB 7024 AGTGAGCGCAACGCCAATTAATGTGAGTTAGTCTACTATTAGGCACCCAGGCTTTACAC 7083
|||||
QY 61 ttatctctccggctcgatgtgtgtgtggaattgtgagcgagataacaatttcacagga 120
|||||
DB 7084 TTTATCTCTCCGGCTCGTATGTTGTGTGAAATTGTGAGCGGATAACAATTTTCACAGGA 7143
|||||
QY 121 aacagctatgaccatgattacgcgaagctcgaataattacccctcaataaagggaacaaaag 180
|||||
DB 7144 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCTCACTAAGGGAACAAAAG 7203
|||||
QY 181 ctggagctcca-cgcgggtggcgcgctctgaaactagtgagatccccccggctgcaggaa 239
|||||
DB 7204 CTGAGCTCCACGCGGTGGCGGCTCTAGAACTAGTGATCCCGGGCTGCAGGAA 7263
|||||
QY 240 ttc 242
DB 7264 TTC 7266

RESULT 8
V14340/c
ID V14340 standard; DNA; 3699 BP.
AC V14340;
DE 19-MAY-1998 (first entry)
DE Plasmid pBSGFP expressing green fluorescent protein.
KW Green fluorescent protein; GFP; blue fluorescent protein; BFP; jellyfish;
KW genetic engineering marker; gene therapy; plasmid; ss.
OS Synthetic.
PN WO9742320-Al.
PD 13-NOV-1997.
PF 07-MAY-1997; U07625.
PR 08-MAY-1996; US-646538.
PA (USSH ) US SEC DEPT HEALTH
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Gaitanaris GA, Pavlakis GN, Stauber NH, Vournakis JN;
WPI: 97-558982/51.
DR New nucleic acid encoding proteins of Aequorea victoria with
PT increased fluorescence - useful as markers for detecting cellular
PT transformation, subcellular localisation of proteins, for assessing
PT gene therapy, mutagenicity etc.
PS Example -: Page 67-68; 105pp; English.
CC This sequence is a plasmid used to express the mutated green fluorescent
CC protein (GFP) of Aequorea victoria coding sequence of the invention. The
CC DNA of the invention that encodes a protein that is (a) a protein that
CC has Leu at position 65 (and optionally Thr at 168 and optionally further
CC Cys at 66) and has cellular fluorescence at least 5 times that of
CC wild-type GFP or (b) is a blue fluorescent protein (BFP) with His at
CC position 67 and also at least one of Leu at 65 and Ala at 164, and has
CC cellular fluorescence at least 5 times that of BFP (Tyr67 to His). The
CC nucleic acids can be used as markers in genetic engineering and gene
CC therapy. They may also be used to detect and characterise regulatory and
CC coding sequence elements that control subcellular expression and
CC targeting of proteins. Typical applications are monitoring targeting and
CC transport of proteins in cells; assessment of gene therapy procedures; in
CC diagnosis (when expressed under control of a promoter induced by a
CC particular analyte); assessment of mutagenicity of compounds; and for
CC drug screening (where expression is controlled by the promoter of a
CC target gene), particularly for antiviral or antiparasitic agents.
CC Sequence 3699 BP; 942 A; 884 C; 898 G; 975 T;
SQ
Query Match 11.0%; Score 230.6; DB 1; Length 3699;
Best Local Similarity 98.0%; Pred. No. 2.1e-41;
Matches 244; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtatgagttagtcactcattagcaccacccaggcttaacac 60

```

```

|||||
Db 1678 AGTAGGCGCAACGCAATTAATGTAGTTAGTCTCACTCATTAGGCAACCCAGCGCTTTACAC 1619
QY 61 ttatgtctcggctcgtatgttgtgtgaaattgtgagcggataacaattttcacacagga 120
|||||
Db 1618 TTTATGCTTCGGCTCGTATGTGTGTGGAAATGTGAGCGGATAACAATTTCCACACAGGA 1559
QY 121 aacagctatgacctgattacgcgaagctcgaaattaaacctcactaaagggaaacaaaag 180
|||||
Db 1558 AACAGCTATGACCATGATTAACGCCAAGCTCGGAAATTAACCTTCATAAAGGGACACAAAG 1499
QY 181 ctgagagctcca-cgcggtggcgcgcgtctagaactagtggtatcccccgggctgcagga 239
|||||
Db 1498 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 1439
QY 240 ttcggttgc 248
|||||
Db 1438 TTCGATCGC 1430
|||||

RESULT 9
ID V13845/c
AC V13845;
DT 16-JUL-1998 (first entry)
DE Complete DNA sequence of plasmid pLF120.
KW Plasmid pLF120; canine adenovirus; vaccine; ss.
OS Synthetic.
PN WO9800166-A1.
PD 08-JAN-1998.
PR 30-JUN-1997; U11486.
PF 03-JUL-1996; US-675566.
PR 03-JUL-1996; US-675556.
PA (INMR ) RHONE MERIEUX INC.
PI Fischer L;
DR WPI; 98-086736/08.
PT Canine adenovirus synthetically modified to contain exogenous DNA -
PT where non-essential region of virus has been deleted, useful in
PT immunogenic, immunological or vaccine composition(s)
PS Example 15; Fig 35; 226pp; English.
CC The present plasmid relates to an invention where a canine
CC adenovirus (CAD) is synthetically modified to contain exogenous DNA,
CC where a non-essential region of the CAD has been deleted.
CC An immunogenic, immunological or vaccine composition comprising the
CC the above CAD can be used to induce an immunological response in a
CC host vertebrate, preferably a canine or human, to which it is
CC administered, or transfer genetic information to an animal or
CC human. The exogenous DNA preferably encodes an expression product
CC comprising an epitope of interest, biological response modulator,
CC growth factor, recognition sequence, therapeutic gene or fusion
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
CC necrosis factor or melanoma associated antigen.
SQ Sequence 6046 BP; 1561 A; 1614 C; 1436 G; 1434 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6046;
Best Local Similarity 99.2%; Pred. NO. 4.3e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtgagttagctcactcattagcaccacagcgtttacac 60
Db 3338 AGTGAGCGCAACGCAATTAATGTAGTTAGTCTCACTCATTAGGCAACCCAGCGCTTTACAC 3279
QY 61 ttatgtctcggctcgtatgttgtgtgaaattgtgagcggataacaattttcacacagga 120
|||||
Db 3278 TTTATGCTTCGGCTCGTATGTGTGTGGAAATGTGAGCGGATAACAATTTCCACACAGGA 3219
QY 121 aacagctatgacctgattacgcgaagctcgaaattaaacctcactaaagggaaacaaaag 180
|||||
Db 3218 AACAGCTATGACCATGATTAACGCCAAGCTCGGAAATTAACCTTCATAAAGGGACACAAAG 3159
QY 181 ctgagagctcca-cgcggtggcgcgcgtctagaactagtggtatcccccgggctgcagga 239
|||||
```

```

Db 3158 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 3099
QY 240 ttc 242
|||||
Db 3098 TTC 3096
|||||

RESULT 10
ID V13844/c
AC V13844;
DT 16-JUL-1998 (first entry)
DE Complete DNA sequence of plasmid pLF100.
KW Plasmid pLF100; canine adenovirus; vaccine; ss.
OS Synthetic.
PN WO9800166-A1.
PD 08-JAN-1998.
PR 30-JUN-1997; U11486.
PF 03-JUL-1996; US-675566.
PR 03-JUL-1996; US-675556.
PA (INMR ) RHONE MERIEUX INC.
PI Fischer L;
DR WPI; 98-086736/08.
PT Canine adenovirus synthetically modified to contain exogenous DNA -
PT where non-essential region of virus has been deleted, useful in
PT immunogenic, immunological or vaccine composition(s)
PS Example 14; Fig 33; 226pp; English.
CC The present plasmid relates to an invention where a canine
CC adenovirus (CAD) is synthetically modified to contain exogenous DNA,
CC where a non-essential region of the CAD has been deleted.
CC An immunogenic, immunological or vaccine composition comprising the
CC the above CAD can be used to induce an immunological response in a
CC host vertebrate, preferably a canine or human, to which it is
CC administered, or transfer genetic information to an animal or
CC human. The exogenous DNA preferably encodes an expression product
CC comprising an epitope of interest, biological response modulator,
CC growth factor, recognition sequence, therapeutic gene or fusion
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
CC necrosis factor or melanoma associated antigen.
SQ Sequence 6245 BP; 1639 A; 1671 C; 1451 G; 1483 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6245;
Best Local Similarity 99.2%; Pred. NO. 4.4e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtgagttagctcactcattagcaccacagcgtttacac 60
Db 3537 AGTGAGCGCAACGCAATTAATGTAGTTAGTCTCACTCATTAGGCAACCCAGCGCTTTACAC 3478
QY 61 ttatgtctcggctcgtatgttgtgtgaaattgtgagcggataacaattttcacacagga 120
Db 3477 TTTATGCTTCGGCTCGTATGTGTGTGGAAATGTGAGCGGATAACAATTTCCACACAGGA 3418
QY 121 aacagctatgacctgattacgcgaagctcgaaattaaacctcactaaagggaaacaaaag 180
|||||
Db 3417 AACAGCTATGACCATGATTAACGCCAAGCTCGGAAATTAACCTTCATAAAGGGACACAAAG 3358
QY 181 ctgagagctcca-cgcggtggcgcgcgtctagaactagtggtatcccccgggctgcagga 239
Db 3357 CTGGAGCTCCACCGCGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 3298
QY 240 ttc 242
|||||
Db 3297 TTC 3295
|||||

RESULT 11
ID V13843/c
AC V13843;
DT 16-JUL-1998 (first entry)
DE Complete DNA sequence of plasmid pLF116.
```

KW Plasmid pLF116; canine adenovirus; vaccine; ss.
 OS Synthetic.
 PN WO9800166-A1.
 PD 08-JAN-1998.
 PF 30-JUN-1997; U11486.
 PR 03-JUL-1996; US-675566.
 PR 03-JUL-1996; US-675556.
 PA (INMR) RHONE MERIEUX INC.
 PI Fischer L;
 DR WPI; 98-086736/08.

PT Canine adenovirus synthetically modified to contain exogenous DNA -
 PT where non-essential region of virus has been deleted, useful in
 PT immunogenic, immunological or vaccine composition(s)
 PS Example 14; Fig 31; 226pp; English.
 CC The present plasmid relates to an invention where a canine
 CC adenovirus (cad) is synthetically modified to contain exogenous DNA,
 CC where a non-essential region of the cad has been deleted.
 CC An immunogenic, immunological or vaccine composition comprising the
 CC the above cad can be used to induce an immunological response in a
 CC host vertebrate, preferably a canine or human, to which it is
 CC administered, or transfer genetic information to an animal or
 CC human. The exogenous DNA preferably encodes an expression product
 CC comprising an epitope of interest, biological response modulator,
 CC growth factor, recognition sequence, therapeutic gene or fusion
 CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
 CC necrosis factor or melanoma associated antigen.
 SQ Sequence 6448 BP; 1740 A; 1698 C; 1482 G; 1527 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6448;
 Best Local Similarity 99.2%; Pred. No. 4.4e-41;
 Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 agtgagcgcaacgcaattaatgtgagttagctcactcattagcaccgccaggtttacac 60
 DB 3740 AGTGAGCGCAACGCAATTAAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 3681
 QY 61 ttatgtctccggcgctgctatgttgtggaattgtgagcgataacaaatttcacagga 120
 DB 3680 TTTATGCTTCCGGCTCGTATGTGTGTGGAATTGTGAGCGGATAACAAATTTACACAGGA 3621
 QY 121 aacagctatgacatgattacgcacagctcgaaattacccctcactaaagggaacaaaag 180
 DB 3620 AACAGCTATGACCATGATTACGCAAGCTCGGAATTAACCCCTCACTAAAGGGAACAAAAG 3561
 QY 181 ctggagctcca-cgcgggtggcgcgctctagaactagtgatcccccggtcgagga 239
 DB 3560 CTGGAGCTCCACCGGGTGGCGCGCTCTAGAACTAGTGTGATCCCCCGGCTGCAGGAA 3501
 QY 240 ttc 242
 DB 3500 TTC 3498

RESULT 12
 V13167/c
 ID V13167 standard; DNA; 6580 BP.
 AC V13167;
 DT 16-JUL-1998 (first entry)
 DE Complete DNA sequence of plasmid pLF086.
 KW Plasmid pLF086; canine adenovirus; vaccine; ss.
 OS Synthetic.
 PN WO9800166-A1.
 PD 08-JAN-1998.
 PF 30-JUN-1997; U11486.
 PR 03-JUL-1996; US-675566.
 PR 03-JUL-1996; US-675556.
 PA (INMR) RHONE MERIEUX INC.
 PI Fischer L;
 DR WPI; 98-086736/08.

PT Canine adenovirus synthetically modified to contain exogenous DNA -
 PT where non-essential region of virus has been deleted, useful in
 PT immunogenic, immunological or vaccine composition(s)

PS Example 6; Fig 7; 226pp; English.
 CC The present plasmid relates to an invention where a canine
 CC adenovirus (cad) is synthetically modified to contain exogenous DNA,
 CC where a non-essential region of the cad has been deleted.
 CC An immunogenic, immunological or vaccine composition comprising the
 CC the above cad can be used to induce an immunological response in a
 CC host vertebrate, preferably a canine or human, to which it is
 CC administered, or transfer genetic information to an animal or
 CC human. The exogenous DNA preferably encodes an expression product
 CC comprising an epitope of interest, biological response modulator,
 CC growth factor, recognition sequence, therapeutic gene or fusion
 CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
 CC necrosis factor or melanoma associated antigen.
 SQ Sequence 6580 BP; 1710 A; 1790 C; 1514 G; 1565 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6580;
 Best Local Similarity 99.2%; Pred. No. 4.4e-41;
 Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 agtgagcgcaacgcaattaatgtgagttagctcactcattagcaccgccaggtttacac 60
 DB 4696 AGTGAGCGCAACGCAATTAAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 4637
 QY 61 ttatgtctccggcgctgctatgttgtggaattgtgagcgataacaaatttcacagga 120
 DB 4636 TTTATGCTTCCGGCTCGTATGTGTGTGGAATTGTGAGCGGATAACAAATTTACACAGGA 4577
 QY 121 aacagctatgacatgattacgcacagctcgaaattacccctcactaaagggaacaaaag 180
 DB 4576 AACAGCTATGACCATGATTACGCAAGCTCGGAATTAACCCCTCACTAAAGGGAACAAAAG 4517
 QY 181 ctggagctcca-cgcgggtggcgcgctctagaactagtgatcccccggtcgagga 239
 DB 4516 CTGGAGCTCCACCGGGTGGCGCGCTCTAGAACTAGTGTGATCCCCCGGCTGCAGGAA 4457
 QY 240 ttc 242
 DB 4456 TTC 4454

RESULT 13
 V13842/c
 ID V13842 standard; DNA; 6613 BP.
 AC V13842;
 DT 16-JUL-1998 (first entry)
 DE Complete DNA sequence of plasmid pLF102.
 KW Plasmid pLF102; canine adenovirus; vaccine; ss.
 OS Synthetic.
 PN WO9800166-A1.
 PD 08-JAN-1998.
 PF 30-JUN-1997; U11486.
 PR 03-JUL-1996; US-675566.
 PR 03-JUL-1996; US-675556.
 PA (INMR) RHONE MERIEUX INC.
 PI Fischer L;
 DR WPI; 98-086736/08.

PT Canine adenovirus synthetically modified to contain exogenous DNA -
 PT where non-essential region of virus has been deleted, useful in
 PT immunogenic, immunological or vaccine composition(s)
 PS Example 13; Fig 29; 226pp; English.
 CC The present plasmid relates to an invention where a canine
 CC adenovirus (cad) is synthetically modified to contain exogenous DNA,
 CC where a non-essential region of the cad has been deleted.
 CC An immunogenic, immunological or vaccine composition comprising the
 CC the above cad can be used to induce an immunological response in a
 CC host vertebrate, preferably a canine or human, to which it is
 CC administered, or transfer genetic information to an animal or
 CC human. The exogenous DNA preferably encodes an expression product
 CC comprising an epitope of interest, biological response modulator,
 CC growth factor, recognition sequence, therapeutic gene or fusion
 CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
 CC necrosis factor or melanoma associated antigen.

```
SQ Sequence 6613 BP; 1724 A; 1781 C; 1531 G; 1576 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6613;
Best Local Similarity 99.2%; Pred. No. 4.4e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtggcgcaacgcaataatgtgagttagctcaactcattagcagcccccaggtttacac 60
DB 3905 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTACCTATAGGACCCAGCGCTTTACAC 3846

QY 61 ttatgtctccggctcgtatgtgtggaattgtgagcggaatacaatttcacacagga 120
DB 3845 TTTATGCTTCGGCTCGTATGTGTGTGGAATTTGTGAGCGGATAACAATTTACACAGGA 3786

QY 121 aacagctatgacatgattacgccaagctcgaaattaaacctcaactaaaggaacaaaag 180
DB 3795 AACAGCTATGACCATGATTAAGCCAGCTCGGAATTAACCCCTCACTAAGGGAACAAG 3726

QY 181 ctggagctcca-cgcggtggcgccgctctagaactagtggtatccccgggctcgagaa 239
DB 3725 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 3666

QY 240 ttc 242
DB 3665 TTC 3663

RESULT 14
Q39050
ID V13165 standard; DNA; 6824 BP.
AC Q39050;
DE 28-JUL-1993 (first entry)
DE K.lactis/S. cerevisiae genetic vector.
KW Genetic; vector; integration; kluyveromyces lactis; 25S ribosomal DNA;
KW Saccharomyces cerevisiae; E. coli; domain; yeast; plasmid; promoter;
KW expression cassette; HIS3; marker; transformant; human; lysozyme; HLZ;
KW GAL7; signal sequence; killer toxin; transcription termination signal;
KW FLP; 2 micron plasmid; ss.
OS Synthetic.
PN EP-537456-A.
PD 21-APR-1993.
PF 31-AUG-1992; 114638.
PR 04-SEP-1991; IT-M12349.
PA (ISTS) SCLAVO SPA.
PI Galeotti CL, Gallo E, Riccio ML, Rossolini GM, Thaller MC;
DR Vector for Kluyveromyces lactis and Saccharomyces cerevisiae -
PT which allows stable multiple integration of DNA for prodn. of
PT heterologous proteins
PS Claim 1; Fig 1; 26pp; English.
CC This sequence represents a genetic vector which allows the stable
CC multiple integration of DNA sequences into the genome of Kluyveromyces
CC lactis and Saccharomyces cerevisiae. This sequence can be used in an
CC integrating vector which comprises a region necessary for the stable
CC maintenance of the plasmid in E. coli and a domain which acts as an
CC integrating unit consisting of two not contiguous sequences of the 25S
CC ribosomal DNA from S. cerevisiae, flanking a genetic marker suitable
CC for selection of the yeast transformants in which the integration
CC event has occurred. Other DNA sequences may be introduced into the
CC integration plasmid, such as expression cassettes. The gene HIS3
CC from K. lactis and S. cerevisiae is pref. used as a genetic marker
CC for the selection of transformants and an expression cassette for the
CC production and secretion into the culture medium of human lysozyme.
CC This complete transformation vector is 7850 bp long and includes the
CC integration vector of the invention and an expression cassette
CC comprising the K. lactis GAL7 promoter, the signal sequence of the K.
CC lactis killer toxin, the cDNA encoding the ripe form of human lysozyme
CC (HLZ) and the transcription termination signal FLP of the 2 micron
CC plasmid from S. cerevisiae.
SQ Sequence 6824 BP; 1815 A; 1521 C; 1726 G; 1762 T;
```

```
Query Match 10.9%; Score 229.4; DB 1; Length 6824;
Best Local Similarity 99.2%; Pred. No. 4.5e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtggcgcaacgcaataatgtgagttagctcaactcattagcagcccccaggtttacac 60
DB 6570 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTACCTATAGGACCCAGCGCTTTACAC 6629

QY 61 ttatgtctccggctcgtatgtgtggaattgtgagcggaatacaatttcacacagga 120
DB 6630 TTTATGCTTCGGCTCGTATGTGTGTGGAATTTGTGAGCGGATAACAATTTACACAGGA 6689

QY 121 aacagctatgacatgattacgccaagctcgaaattaaacctcaactaaaggaacaaaag 180
DB 6690 AACAGCTATGACCATGATTAAGCCAGCTCGGAATTAACCCCTCACTAAGGGAACAAG 6749

QY 181 ctggagctcca-cgcggtggcgccgctctagaactagtggtatccccgggctcgagaa 239
DB 6750 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGGATCCCGCGGCTGCAGGAA 6809

QY 240 ttc 242
DB 6810 TTC 6812

RESULT 15
V13165/C
ID V13165 standard; DNA; 6959 BP.
AC V13165;
DE 16-JUL-1998 (first entry)
DE Complete DNA sequence of plasmid pLF047A.
KW Plasmid pLF047A; canine adenovirus; vaccine; ss.
OS Synthetic.
PN WO9800166-A1.
PD 08-JAN-1998.
PF 30-JUN-1997; U11486.
PR 03-JUL-1996; US-675566.
PR 03-JUL-1996; US-675556.
PA (INMR) RHONE MERIEUX INC.
PI Fischer L;
DR WPI; 98-086736/08.
PT Canine adenovirus synthetically modified to contain exogenous DNA
PT where non-essential region of virus has been deleted, useful in
PT immunogenic, immunological or vaccine composition(s)
PS Example 6; Fig 3; 226pp; English.
CC The present plasmid relates to an invention where a canine
CC adenovirus (CAD) is synthetically modified to contain exogenous DNA,
CC where a non-essential region of the CAD has been deleted.
CC An immunogenic, immunological or vaccine composition comprising the
CC the above CAD can be used to induce an immunological response in a
CC host vertebrate, preferably a canine or human, to which it is
CC administered, or transfer genetic information to an animal or
CC human. The exogenous DNA preferably encodes an expression product
CC comprising an epitope of interest, biological response modulator,
CC growth factor, recognition sequence, therapeutic gene or fusion
CC protein, e.g. a Morbillivirus antigen, rabies glycoprotein, tumour
CC necrosis factor or melanoma associated antigen
CC Sequence 6959 BP; 1834 A; 1880 C; 1591 G; 1653 T;

Query Match 10.9%; Score 229.4; DB 1; Length 6959;
Best Local Similarity 99.2%; Pred. No. 4.5e-41;
Matches 241; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 agtggcgcaacgcaataatgtgagttagctcaactcattagcagcccccaggtttacac 60
DB 4264 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTACCTATAGGACCCAGCGCTTTACAC 4205

QY 61 ttatgtctccggctcgtatgtgtggaattgtgagcggaatacaatttcacacagga 120
DB 4204 TTTATGCTTCGGCTCGTATGTGTGTGGAATTTGTGAGCGGATAACAATTTACACAGGA 4145

QY 121 aacagctatgacatgattacgccaagctcgaaattaaacctcaactaaaggaacaaaag 180
```

|||||
Db 4144 AACAGCTATGACCATGATTAGCCCAAGCTCGGAATTAAACCTCACTAAAGGGAAACAAAG 4085
QY 181 ctggagctcca-cgcggtgcccgcgcctctagaactagtagtatccccccgggctgcaggaa 239
Db 4084 CTGGAGCTCCACCGCGGTGGCGCGCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA 4025
QY 240 ttc 242
Db 4024 TTC 4022

Search completed: August 1, 2000, 18:20:53
Job time: 20174 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 1, 2000, 13:36:13 ; Search time 9137.45 Seconds
(without alignments)
411.142 Million cell updates/sec

Title: US-09-147-955-11
Perfect score: 2105

Sequence: 1 agtgcgcgaacgcaattaa.....acaaaaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pil.*
8: gb_p12.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: em_fun.*
17: em_hum1.*
18: em_hum2.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_sy.*
29: em_un.*
30: em_v1.*
31: gb_htg1.*
32: gb_htg2.*
33: gb_in1.*
34: gb_in2.*
35: em_ba1.*
36: em_ba2.*
37: em_hum3.*
38: em_hum4.*
39: gb_pr4.*
40: gb_htg3.*
41: gb_htg4.*
42: gb_htg5.*
43: gb_htg6.*

44: gb_htg7.*
45: em_htg1.*
46: em_htg2.*
47: em_htg3.*
48: em_hum5.*
49: gb_pl3.*
50: gb_pr5.*
51: gb_htg8.*
52: gb_htg9.*
53: gb_htg10.*
54: gb_htg11.*
55: gb_htg12.*
56: gb_htg13.*
57: gb_htg14.*
58: gb_in3.*
59: gb_htg15.*
60: gb_htg16.*
61: gb_htg17.*
62: em_htg4.*
63: em_htg5.*
64: em_htg6.*
65: em_htg7.*
66: em_hum6.*
67: gb_htg18.*
68: gb_htg19.*
69: gb_htg20.*
70: gb_htg21.*
71: gb_htg22.*
72: gb_htg23.*
73: gb_htg24.*
74: gb_htg25.*
75: gb_htg26.*
76: gb_htg27.*
77: gb_htg28.*
78: gb_htg29.*
79: gb_htg30.*
80: gb_htg31.*
81: gb_v11.*
82: gb_v12.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1864	88.6	1864	7	AB027455
2	455.4	21.6	1507	7	AB013596
3	455.4	21.6	1594	7	AB000623
4	439.8	20.9	1458	7	AB013597
5	424.2	20.2	2043	7	AB013598
6	303.2	14.4	19894	8	ATCHRIV38
7	303.2	14.4	200576	7	ATFCA0
8	263	12.5	195068	8	ATCHRIV42
9	263	12.5	197419	8	ATCHRIV41
10	263	12.5	205065	7	ATFCA4
11	259.8	12.3	1589	8	ATU81293
12	241.4	11.5	84203	49	AC005106
13	241.4	11.5	103223	8	AC007153
14	232.8	11.1	3481	12	MARIREDM2
15	232	11.0	476	1	CSPAG13
16	231	11.0	2958	14	ARBLSKM
17	231	11.0	2958	14	ARBLSKP
18	231	11.0	3062	14	CVU47947
19	231	11.0	5192	14	PVPJ2581
20	230.4	10.9	1353	5	A67220
21	229.4	10.9	2964	14	SYNBLSKP
22	229.4	10.9	2964	14	SYNBLSKV
23	229.4	10.9	6824	5	A25909
24	229.4	10.9	7372	14	SYNPHSCSKV
					AB027455 Petunia x
					AB013596 Perilla f
					AB000623 Nicotiana
					AB013597 Perilla f
					AB013598 Verbena h
					AL161538 Arabidops
					297335 Arabidops
					AL161542 Arabidops
					AL161541 Arabidops
					297339 Arabidops
					U81293 Arabidops
					AC005106 Genomic s
					AC007153 Arabidops
					X68127 M.auratus m
					Z47172 Calothrix D
					X5324 pBluescript
					X5325 pBluescript
					U47947 Cloning vec
					X98363 Cloning vec
					A67220 Sequence 4
					L08787 BlueScribe
					L08786 BlueScribe
					A25909 Synthetic y
					L08874 PhageScript

C 25	228.2	10.8	4133	14	AF106619	AF106619 Cloning v
C 26	227.8	10.8	1832	10	D34613S02	D34614 Human TBXAS
C 27	227.8	10.8	2961	14	ARBL2SKM	X52330 paluescript
C 28	227.8	10.8	2961	14	ARBL2SKM	X52328 paluescript
C 29	227.8	10.8	2976	14	CVU611229	X61229 Cloning vec
C 30	227.8	10.8	3633	14	ASPGREEN2	Y09374 Artificial
C 31	227.8	10.8	3886	14	XU035129	XU3129 Plasmid pBS
C 32	227.8	10.8	4051	14	XU035126	XU3126 Plasmid pBS
C 33	227.8	10.8	4285	14	XU035137	XU3137 Plasmid pBS
C 34	227.8	10.8	4480	14	XU035134	XU3134 Plasmid pBS
C 35	227.8	10.8	4669	14	XU035130	XU3130 Plasmid pBS
C 36	227.8	10.8	4755	14	U937113	U93713 Cloning vec
C 37	227.8	10.8	4842	14	XU351125	XU3125 Plasmid pBS
C 38	227.8	10.8	4893	14	U937114	U93714 Cloning vec
C 39	227.8	10.8	5035	14	XU035128	XU3128 Plasmid pBS
C 40	227.8	10.8	5060	14	XU035127	XU3127 Plasmid pBS
C 41	227.8	10.8	5275	14	XU035135	XU3135 Plasmid pBS
C 42	227.8	10.8	5407	14	U937115	U93715 Cloning vec
C 43	227.8	10.8	5434	14	AF016889	AF016889 Cloning v
C 44	227.8	10.8	5528	14	U937117	U93717 Cloning vec
C 45	227.8	10.8	5638	14	CVU14120	U14120 Cloning vec

ALIGNMENTS

RESULT	1
AB027455	
LOCUS	1864 bp mRNA
DEFINITION	Petunia x hybrida PH1 mRNA for anthocyanin 5-O-glucosyltransferase; complete cds.
ACCESSION	AB027455
VERSION	AB027455.1 GI:6683051
KEYWORDS	anthocyanin 5-O-glucosyltransferase.
SOURCE	Petunia x hybrida (cultivar:Surfinia) corolla cdNA to mrna.
ORGANISM	Petunia x hybrida
REFERENCE	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; angiosperms; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
AUTHORS	1 (bases 1 to 1864)
TITLE	Yamazaki,M. and Saito,K.
JOURNAL	Anthocyanin 5-O-glucosyltransferase
REFERENCE	Published Only in Database (2000) In press
AUTHORS	2 (bases 1 to 1864)
TITLE	Yamazaki,M. and Saito,K.
JOURNAL	Direct Submission
NOTE	Submitted (17-MAY-1999) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Chiba University, Faculty of Pharmaceutical Sciences; Inage-ku Yayoi-cho 1-33, Chiba, Chiba 263-8522, Japan (E-mail:mamiyep.chiba-u.ac.jp, Tel:81-43-290-2905, Fax:81-43-290-2905)

BASE COUNT	603 a	318 c	369 g	574 t
ORIGIN	GSSHKNLKAFIDVAKGF*			
Query Match	88.6%	Score 1864;	DB 7; Length 1864;	
Best Local Similarity	100.0%	Pred. NO. 0;		
Matches 1864; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
QY	242	cgggtgctgcagacaaatttcacaaaaccagaataatgaacctcccttcccccccctaaaa	301	
Db	1	CGGTTCGTCTGCCACAAATTTACAAAACAAGAATAAAGCATCCCTTCCCCCCTTAAA	60	
QY	302	aaacatacaagttttaattttcactcaagcagaasaaatatggtgcagcctcatgcac	361	
Db	61	AAACATACAAGTTTTTAATTTTTTTCACTAAGCAGAANAATATSGTGACGCTCATGTGCATC	120	
QY	362	ttaacaacatttcagacacaaaggccaatatataaccagacacttcaatttgcacaagaattott	421	
Db	121	TTAACAAACATTTCGAGCACAAAGGCCATTATATTCAGACACTTCAATTTGCCAAGAANTCTT	180	
QY	422	gtcaagatggcgatagaagtgcacattttctacaagcatttatgcccaagcgcgtatgat	481	
Db	181	GTCGAAGATGGGCATAGAAGTGACATTTCCTACAAGCATTTATGCCCCAAGCGGTATGGAT	240	
QY	482	gaaaaatccattctaattgcaccaaaaggattgaatttcatttccatttccagtggcttt	541	
Db	241	GAAAAATGCATCTTTATTCACCANNAAGGATTGAATTTTCATTTCCTATTTCCCGATGGCTTT	300	
QY	542	gatgaaggttttgatcatccaagaagaccctgtattttacatgtccacaacttcgtaaatgt	601	
Db	301	GATGAAGGTTTTTGATCATTTCAAAGAGACCCCTGTATTTTACATGTCAACAACCTCGTAAATGT	360	
QY	602	ggaagtgaacctgtcaaaaaataattctcacctgcctcgtgaaaatggacagcctataact	661	
Db	361	GGAAGTGAACCTGTCAAAAAATAATTCCTCACTTGCTCTGAAATGGACAGCCTATAACT	420	
QY	662	tgcttaacttactocattttccctcttggggacagagaggtgacagtgaaagttcaaatc	721	
Db	421	TGCCATACTTTATTCATTTTCCCTTTCCTTTGGCGACAGAGGTAGCAGTGAAGTTCACATC	480	
QY	722	cctctgcctctcttggagtcacaccagcaaacatttggacatatattacttcaaacttt	781	
Db	481	CCTTCTGCTCTCTTTGGAGTCAACACGACAACAAATATTTGCATATATATTACTTCAACTTT	540	
QY	782	catgatataaaaagctatgcttaatgaatccaatgataccaaatggcccaactcaact	841	
Db	541	CATGATATATGAAAAGCTATGGCTTAATGAATCCAATGATCCAAATTTGGTCCATTCACATT	600	
QY	- 842	ccggggcttccactactygaactcgagatctctcttcatttttacttcccttatgggtca	901	
Db	601	CCGGGGCTTCCACTACTGGAACCTCGAGATCTCTCTTCATTTTACTTCTTATGGTGCA	660	
QY	902	aaaggagctctcaggtgcacttccaccattcaagaattatgacacatttagatgct	961	
Db	661	AANGGAGTCTTCGAGTTGCACTTCCACCAYTTCANNANATTTGNAGACNATTAGATGCT	720	
QY	962	gaaacacactctaagattcttgtgaaatacatatttgatgaattagagccctgagcactcaat	1021	
Db	721	GAACACACTCTCTAAGATTCTTGTGAATACATTTTGATGAATTAGAGCCTGAGGACATCAAT	780	
QY	1022	gcaattgaaggtataaagtttataagtttggaccgttgatctctctctctctcttctgggt	1081	
Db	781	GCAATTGAAGGTATATAAGTTTATGGAAATGGACCGTGTATTCCTCTCTCTTCTTGGGT	840	
QY	1082	ggaatgacctcttagatgcttcaatttgggtgggtgactctttttcaaaattcaaatgactat	1141	
Db	841	GGAAATGACCCCTTTAGATGCTTCATTTTGGTGGTGATCTTTTCAAAATTTCAANTGACTAT	900	
QY	1142	atgggaatgggttaaactcaagcnaaatctcatcagtttgtttatatcatcttttggggagtcta	1201	
Db	901	ATGGGAATGGTTAAACTCAAGCCAAATTCATCAGTGTGTTATATATCTTTTGGGAGTCTA	960	

QY	1202	atgaatccatctattagccaaataggagagatatcaaaagggttagagacatagggaagg	1261
Db	961	ATGAATCCATCTATTAGCCAAATAGGAGAGATATCAAAAGGGTTGATAGCATAGGAAGG	1020
QY	1262	ccgtttttatgggtgataaaagaataatgaaaaaggccaaagaagagaataaaaaagctt	1321
Db	1021	CCGTTTTTATGGGTGATAAAAGAAATGAAAGAAAGCCAAAGCAAGAGAGATAAAAGCCT	1080
QY	1322	ggttgtattggaagaattggaataataggaaaaataagttccatgggtgtccacaactgaa	1381
Db	1081	GGTGTATTGGAAGATTGGAAAAATAGGAAAAATAGTCCATGGGTGTCACTAGCTTGAA	1140
QY	1382	gttctaaacacatccatctttagtagtctttgttcttcattgttgatgggaattccagctcta	1441
Db	1141	GTTCCTAAAAACATCCATCTTTAGGATGTTTTGTCTTCATTGTGGATGGAAATTCAGCCTTA	1200
QY	1442	ggagcttagctgtggagtgccagttgtgcattctcctaattgacagagatcaaatgaca	1501
Db	1201	GAGAGTTTAGCTGTGGAGTCCAGCTTGTGGCATTTCCTCAATGGACATCAATGACA	1260
QY	1502	aatgcacacaagttgaagatgtgtggaaaagtggagtaagatgagagataaaatgaagat	1561
Db	1261	AATGCCAAACAAGTTGAAGATGTGTGGAAGTAGTGGAGTAAAGCTAGAGATGAGATAAATGAAGAT	1320
QY	1562	gctgtgtttaaagtagaagaatcaaaagggttatctgaattgcatgatatgagtaggagagag	1621
Db	1321	GGTGTGTTGAAGTGAAGAAATCAAAAGGTGTATTGTAATGTAATGATGGAGAGG	1380
QY	1622	aaaggggaagattgagaagaatgctcaagaaaatggaagaaattggctagagaagctgtg	1681
Db	1381	AAAGGGGAAGAAATTTGACAAGAATGCTAAGAAATGGAAGAAATTTGCTTAGAGAAGCTGTG	1440
QY	1682	aagggaagctgcatctcacacaagaatttaaaggcttttattgatgatgttgccaaaagg	1741
Db	1441	AAGGAAGTGGATCTTCACACAAGAAATTAAGAGCTTTTATTGATGATGTTGCCAAGGG	1500
QY	1742	ttttaattttacaggcttttgcgtgatataacttcccccttagttggcgattcaactcttt	1801
Db	1501	TTTTAATATTTACAGGCTTTTGCCGTGATATTACTTCCCTAGTTGGCGAATTCACCTTT	1560
QY	1802	gtggacttgcttgacaaaaaacctgagggaatgtctgaagacagctaaatgctttaagaag	1861
Db	1561	GTGGACTTGCTTGACAAAANAACTAGGGNAATGCTCTAAGCACAGCTTAATGCTTTAAGNAG	1620
QY	1862	tcatttccaagcttgagcgtgcttttaaactattatccagtaatactatagggttct	1921
Db	1621	TCATTTCCAAGGCTTGAGCGCTCTTTTAAAACTATTAGCCAGTAATCTATAGGGTTCT	1680
QY	1922	ctctattttctctgctctcttttttagccttttcttccaaaggtttaagaatagcgt	1981
Db	1681	CTTCTATTTTCTCTCTCTCTTTTAGCCCTTTTCTTTCCAAGGTTTAAAGATAGCGT	1740
QY	1982	gaacatgcttagtaogtagcttggtatctctatctattaccaagtgcaagattatgctta	2041
Db	1741	GAACATAGCTTAGTACGTAGTCTTGATCTCTATCTTACCAAGTCCAAGATTATGCTTA	1800
QY	2042	tgctgctctctaaattctctaataatgcaagatgaaagaatgacaaaaaataaaaaa	2101
Db	1801	TGCTGTCTCTCTAAATTTCTTAATAAATGCAAGATGAAAAGTACAAAAAANAANA	1860
QY	2102	aaaa 2105	
Db	1861	AAAA 1864	
RESULT	2		
LOCUS	AB013596		
DEFINITION	AB013596 1507 bp mRNA	PLN	20-MAR-1999
ACCESSION	Perilla frutescens PF3R4 mRNA for UDP-glucose:anthocynln		
VERSION	5-O-glucosyltransferase, complete cds.		
KEYWORDS	AB013596		
	AB013596.1 GI:4115558		
	UDP-glucose:anthocynsln 5-O-glucosyltransferase.		

```
QY 631 cacttgcttgaaatggagagcctataacttgctacttttactccatttcccttctcttg 690
Db 313 C-----AACACACGACGCTCAGTTCGTCGTCTACTCCCACTCTTTGTCATG 360
QY 691 ggcgcagaggtagcagctgaagtccacatccctctctctctctctctctctctctctct 750
Db 361 GCGCGGAGGTGGCGGTGAGTCCCGAGGTCCCGAGCGCCCTTCTCTGGGTGAGCCGCG 420
QY 751 aacaaatttgacatatattacttcaacttctatgatgatgaataagctatgcttaataga 810
Db 421 CACCGTGTGTGCTATATTACTTCTCAACGGCTACGACGAGATCGACGCGCG 480
QY 811 atcacaatgatcaaatggctccatccaaactcccggtctccactactcggaaactcgaga 870
Db 481 TTCGACGAA-----ATTGAGCTCCCTCGCTCCACCCCTGGACGCGAG 528
QY 871 tcttctctcatcttctctctctctctctctctctctctctctctctctctctctctcc 930
Db 529 TCTTCGACCTTCTCTGCTGCC---GGAGACACCGGAGAGATTCGCGTGTGATGATG---- 580
QY 931 attcaagaattgatagacacattagctgctgaaccactcctaaactcctgaattctctgaaat 990
Db 581 ----AAGGAGAAGCTGCAAACTTAGACGGTGAAGAGAGAGCGAAAGTGTGGTGAACAC 636
QY 991 attgatgaattagagcctgagcactcaatgcaattgaagtgattataagttttatggaat 1050
Db 637 GTTTGATCGGTGGAGCCGATGCACATCGGCTATTGATAGTATGATGATGATGATGATGAT 696
QY 1051 tggaccgttgattctctctctctctctctctctctctctctctctctctctctctctct 1110
Db 697 CGGCGCTGTGATTCCTTCGCTCTCTGGACGGCGGAGATCCCTCCGAAACGCTCTACGG 756
QY 1111 tgggatctttttcaaatcaaatgactata-----tggaaatggttaactcaagacc 1164
Db 757 CGCGCATCTTTTCGAAATAATCGGAGGAGAAATAACTCGCGTGGAGTGGTGGACAGCAAGCC 816
QY 1165 aaattcatcagttgtttatatctcttcttggagctctaataatgaatcctctattagccaaat 1224
Db 817 GAAATCTTCGCTGTGTATGTCGTCTGGAGCGCTTTTGGAGTTTCCAAAGGCACAAAT 876
QY 1225 ggaaggaatatcaaaaggttgatagatagatgaagggcgtttttatttggtgataaaga 1284
Db 877 GGAAGAGATGGGAAGAGGCTATTAGCTCGCGAAGCGCGTCTTTTATGATGATGATGATGATG 936
QY 1285 aaatgaanaagccaa---agaagaagagataaaaaagcttggttattgaagaattgga 1341
Db 937 ACAGAAGAATGACGACGCGGAAGAAGAAGAAGAGAGTTGAGTTGCATTTGGGAATTCAA 996
QY 1342 aaaaataggaataatagttccatggtgttcacaaacttgaaactgaagttcctcaaacatcttt 1401
Db 997 AAAAATGGGGAATAATGTTCTGCTGCTCGCGATTTGGAGGTTCTGGCGCACCTCGCTT 1056
QY 1402 aggatgttttctctctctctctctctctctctctctctctctctctctctctctctct 1461
Db 1057 GGGATGTTTCTGACGCTTGTGGTGGAACTCGGCTGGAGAGCTTGGAGTGTGAGTTCGGGGT 1116
QY 1462 gccagttgtgcatcttctcaatggacagatcaaatgacaaatgcaaacaaagttgaaga 1521
Db 1117 TCCGCTGTGGCGTGGCGCGCTGCTGTTGATCATCAGACGCAATGCAAGCTGATGTAGGA 1176
QY 1522 tgttggaagtggaagtaagagtgagaataaatgaagaatggtgttggtgaaagtgaaga 1581
Db 1177 TCGCTGGGGACAGCGGTGAGATGAGATGAATGAATGAATGAATGAATGAATGAATGAATGA 1236
QY 1582 aatcaaaaggtgtatgaattgtatgtatgaggaaggaaggaaggggaagattgaagaa 1641
Db 1237 CATGACAGAGGTGTCTGGAGATGTTGATGATGGGTGGGTGGAGAGCAAACTAGTGAGAGA 1296
QY 1642 gaatgcaagaataatggaagaatttggtctagagaagctgtggaagggaagtggtcttccaa 1701
Db 1297 AAATGCCATAAATGGAAGACTTTGGCCAGAGAACGATGGGAGAGGATGGATCTTCACT 1356
```

```
QY 1702 caagaattaaaggctttttatttgatgatgttc 1734
Db 1357 CAAGAATCTCAAGCGCTTTCTTCTCATCAAGTTC 1389

RESULT 3
LOCUS AB000623 1594 bp mRNA PLN 05-FEB-1999
DEFINITION Nicotiana tabacum mRNA for glucosyl transferase, complete cds.
ACCESSION AB000623
VERSION AB000623.1 GI:18053358
KEYWORDS JIGT; glucosyl transferase.
SOURCE Nicotiana tabacum (strain:BY-2) suspension culture cDNA to mRNA.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Solanales; Solanaceae; Nicotiana.
KoJima,H.
Direct Submission
Submitted (23-JAN-1997) to the DDBJ/EMBL/GenBank databases. Hisae
KoJima, school of agriculture,nagoya university, laboratory of
biochemistry, nagoya, furo-cho,chikusa 464-01, Japan
(E-mail:hisae@nuagr1.agr.nagoya-u.ac.jp, Tel:052-789-4098,
Fax:052-789-4094)
2 (bases 1 to 1594)
REFERENCE
AUTHORS KoJima,H., Hashizume,K., Imanishi,S. and Nakamura,K.
TITLE Jasmonate-induced potential glucosyltransferase from tobacco
suspension cell
JOURNAL Unpublished (1997)
FEATURES
Location/Qualifiers
source 1..1594
/organism="Nicotiana tabacum"
/strain="BY-2"
/db_xref="taxon:4097"
/tissue_type="suspension culture"
52..1455
/gene="JIGT"
52..1455
/gene="JIGT"
/standard_name="jasmonate-induced potential
glucosyltransferase"
/codon_start=1
/product="glucosyl transferase"
/protein_id="BAAL9155.1"
/db_xref="GI:1805359"
/translation="MENLKNECHLVIALFPQGCHINPSLOPSKKLINLGVKVLSSSL
SAFNRIKLPKIEGLTFAPFSDGYDGNKGSFDVYHFNLSAIKSHGSEFIANLIKSKA
KNYPTRTVIYTLMDWAGSAVKLHPISTLFWQATVPDIYYRTNPNFYKVD
SODLIELGLPSLSSSDPSPVDVKSNDWAVESIKRQIELLNSENPRILVNTFD
ALENALRVLKNVYMGVGLIPSPFLDEKDKRKNFAADMIENSNMELDARANK
SVIYIFGSAETSISSQWMEELISQGLKCGRPFLWIRETLNGERPEKRLCTCKDELEKI
GRIVRCSQMEVLKSHSSVCFTHCGWNSLTSASGVIIVACPIWNDQICNAKLQD
VWIKGVYRVNANKEGIIRDFQKCIETVMDAERGEELRKNQAOKWKLAKESTKENS
SNYLNKAYVNECLLGH"
1594
polyA_site /note="35 A nucleotides"
BASE COUNT 514 a 238 c 325 g 517 t
ORIGIN

Query Match 21.6%; Score 455.4; DB 7; Length 1594;
Best Local Similarity 60.8%; Pred. No. 4.8e-81;
Matches 839; Conservative 0; Mismatches 516; Indels 24; Gaps 5;

QY 353 catgtcatcttaacaacatttcacagcagaaggccatttaaccagccttcatttgc 412
Db 76 CATGTCCTTATTGTCATTATTCCAGCGCAAGGTCATTAAATCCATCCCTCAATTTTC 135
QY 413 aagaattctgtcagaatggcgatagagtgacattttctcacaagcattatgccccaaagc 472
Db 136 AAAAAATTAATCAATTTAGTGTGTAAGTGTGAAAGTCACCTTTATCTTCTTATCCTATCAGC 188
```

```
QY 473 cgtatggatgaaaaatccattcttaataagcaccacaaagaggattgaatttcattccattttcc 532
Db 189 --TTTCAATCGCATAAAAATCTCCCAAAATAGAGGATTAACCTTCTCTCTCTCTCTCTCTCT 246
QY 533 gatggctttgatgaaggttttgatcattca---aaagacctgtattttacatgtcacaa 589
Db 247 GATGGCTATGATGGAATTTTCAAGGTTTCATTTGATGATTAACCATTTATTTCRAATTCAGCA 306
QY 590 cttcgtataatggaagtgaaactgcaaaaaataattcactgcctgctgaaaaatgga 649
Db 307 ATAAATCTCATGGTTCGAATTTATTGCCCAATTTAATTAATCCAAAGGCAAAAATGGT 366
QY 650 cagcctataacttgctacttacttactccatttctctctctgggcagcagaggttagcacgt 709
Db 367 TACCCCTTTACACGCTGTTATTATACATATTCTTATGGATTGGGCTGTTTCAGTAGCAAAA 426
QY 710 gaagtccaatccctctgcctctctctcttgagtcacacagcaacaatatatggacatat 769
Db 427 AATATACATATCCCTTACATATTTTGGATTCAACGACTACAGCTTTTGTGATATTTAT 486
QY 770 tactcaactttcatggatgaaaaagctatggctgaatgaatccaatgatccaaattgg 829
Db 487 TATTATCGGTTTACTAATTTTGCATAATTTATTTCAGAATATG---ATTCTCAGGATCAA 543
QY 830 tcaattcaacttccggctctccactactggaactgagatcttctcctcatttcttactt 889
Db 544 ATTATTGAGCTACCAAGGTTTCCCTGCTACTTAGTTCAAGTAGTATTCTCTCATTTGTTGTT 603
QY 890 ccttatgggtcaaaagggagctctcagcttgcacttccacttcaaaagaattgatagac 949
Db 604 GATGATGTGAA-----CTCTAATGATTTGGSCAGTTGAGTCAATTAAGGCAAAATTGAG 657
QY 950 acattagatgctgaacacactccttaagattcttctgtaatacatttgaatagtagacct 1009
Db 658 ATATTGAATAGTGAGAAATTCAGAAATCTTGTGTAATCTTGTGATCTTTGATGATTGGAAT 717
QY 1010 gaggcactcaatgcgaatgaaggtttaagtttttatggaattggacctgtgattcctct 1069
Db 718 AATGCTTTTTCAGGGTTTGAATAATGTAATCTATGTTGGAATGAGCTTTAATTTCTCTCA 777
QY 1070 gcttctctgggtggaatgaccttctagatcttcatcttctggtgattcttcttcaaat 1129
Db 778 TCTTTTCTTGTGAGAGAGATGGAAGAATTAATTTTTCGGCTGATATGATCGAAAGT 837
QY 1130 tcaaatgaactataggaatgggttaaaactcaagcgaatccatccatcagttgtttatatct 1189
Db 838 GAAAAATAATATACATGAGTGGTGTGATGCAAGGGCGAATAAATCAGTTATTACATAGCA 897
QY 1190 ttggcagctcaatgaatccatctattagccaaatgagagagatatcaaaaggttgata 1249
Db 898 TTTGGTAGCTATGTCGATGATCAAGTCATATGGATGGAAGAGATTAGTCAAGGATTGTTG 957
QY 1250 gacataggaagccgttttttatgggtgataaaagaaaaatgaaaaagcaagaagaag 1309
Db 958 AAATGTGGAAGACCAATTTTGTGGTGATTAGAGAACCCCTAAATGGGGAACCCGAG 1017
QY 1310 aataaaagcttggttgatgaagaattgtaaaataagaaaaatagtaattcatggtgt 1369
Db 1018 ---GAAAGTGTACTTGTAAAGATGAATTTGGAAGATAGGAGATAGTGCCTTGGTGC 1074
QY 1370 tcacaaatgaagttcttaaaacatccatttaggatgtttgtttctcatttgatgg 1429
Db 1075 TCGCAAAATGGAAGTCTAAACACACTGCTCCGTAGCGTGTTTCTTGACTCATTTGGGATG 1134
QY 1430 aattgccttgagagatttagcttgtagtgcagttgtagcattctcctcaatggaca 1489
Db 1135 AATTGACACTCTGGAGAGCTTAGCTTCTGGTGTGCAATCGTGGCATGCCGATTTGGAAAT 1194
QY 1490 gatcaaatgacaaatgcaaaccaagttagaattgtggaagaagtgagtaagtgaga 1549
Db 1195 GATCAAAATTTGCAACCGAAGCTCATCTCAAGATGTTTGGAGATGTTGGTGAGAGTGAAC 1254
QY 1550 ataaatgaagatggtgtgttgtaagtgagggaatacaaaaggtgtattgtaattggtta 1609
```

```
Db 1255 GCTAATAAAGAGAGGTATCATTTAAAGAGACGAGTTCCAGAGTGTATAGAGATCGTAATG 1314
QY 1610 gatggaggagaaagggggaagaaattggaagaagaaattgtaagaagaattggaattggct 1669
Db 1315 GGAGATGCTGAGGAGGAGAGGAATTTGAGAAAGAAATGCCCAAAATGAAGGATTGGCT 1374
QY 1670 agagaagctgtaaagaaggtggtatcttcacacaagaatttaaggcttttatgatga 1728
Db 1375 AGGAATCTACGAAGAAATAGTTCTATCTAATGTAAATCTTAAGGCTTATGTTAATGA 1433

RESULT 4
AB013597
LOCUS AB013597 1458 bp mRNA PLN 20-MAR-1999
DEFINITION Perilla frutescens PF3R6 mRNA for UDP-glucose:anthocyanin
5-O-glucosyltransferase homologue, complete cds.
ACCESSION AB013597
VERSION AB013597.1 GI:4115560
KEYWORDS PF3R6; UDP-glucose:anthocyanin 5-O-glucosyltransferase homologue.
SOURCE Perilla frutescens (variety:crispa, cultivar:Shikun) leaf cDNA to
mRNA.
ORGANISM Perilla frutescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Gentianales; Lamiales; Lamiaceae; Perilla.
REFERENCE 1 (sites)
AUTHORS Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y.,
Kusumi,T. and Saito,K.
TITLE Molecular cloning and biochemical characterization of a novel
anthocyanin 5-O-glucosyltransferase by mRNA differential display
for plant forms regarding anthocyanin
J. Biol. Chem. 274 (11), 7405-7411 (1999)
99167509
REFERENCE 2 (bases 1 to 1458)
AUTHORS Yamazaki,M., Saito,K. and Gong,Z.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami
Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. &
Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan
(E-mail:maniy@p.chiba-u.ac.jp, tel:+81-43-290-2905,
Fax:+81-43-290-2905)
FEATURES
source
1..1458
location/Qualifiers
/organism="Perilla frutescens"
/cultivar="Shikun"
/variety="crispa"
/db_xref="taxon:48386"
/tissue_type="leaf"
29..1360
/gene="PF3R6"
29..1360
/gene="PF3R6"
/codon_start=1
/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase
homologue"
/protein_id="BAA36422.1"
/db_xref="GI:4115561"
/translation="MVRRLVLLATFPAGHINPALQFAKRLKAGTDVTFSTVYAWR
RMAATSAAGNPFGLDFVAFSDGYDDGLKPGDGKRYSEMKAARGSEALRNLINQ
DVTFFVYSHLFAWAEVARLSHPVTLVLEVPATVLCIHFYFNGYADIDAGSNEIQ
LPRLSEORSPLPTFLPATPERERLMMKEKLETLDEEKAKVLVNTFDALPDLATA
IDRYELIGTLPISATLDGDSYSGDLPEKSENNCNEVLSNPKPSVYVVF
GSYLPRPQAOMEIIGKLLAGGRPFLMWIREQNDGDEEESEELSCIGLKKWKI
SWNSQLEVLAHAPCFVTHCGWNSAVESLSCGIPVAVPQFDDTTNAKLIEDAWG
TCVRYRMNBBGGVGDCEIERCEVMYNDGGDKTLVRENAIKWTKLARQAG"
polyA_site
/note="16 a nucleotides"
BASE COUNT 353 a 348 c 438 g 319 t
ORIGIN
```

Query Match 20.9%; Score 439.8; DB 7; Length 1458;

Best Local Similarity 61.0%; Pred. NO. 6.3e-78; Matches 865; Conservative 0; Mismatches 497;					7;
QY	337	aaatagtgagcgcctcatgtcattcttaacacatttcacagacagccatattatcc	396		
Db	25	AAAATGTCGCGCGCGCTGCTGTAGCAACGTTTCGCGCAAGGCCACATAATCC	84		
QY	397	agcacttcaatttgcagaatcttgcagaatggcagatagaagtcacattttcacaag	456		
Db	85	CGCCCTCCANTTCGCCAAGAGATCTTAAGCGCGCATGACGTCACTGTTTTCACGAG	144		
QY	457	cattatgcacaaagccgtatgatgaataatccattcttaatgc-----accaaagg	510		
Db	145	CCTTATGTCATGGCGCGCATGCGCAACACAGCCTCGCGCTCGCGGAACCCACCGG	204		
QY	511	attgaattcattccatttccatagcgttttaataaggttttgatcattcaaaagacc	570		
Db	205	CCTGCACTTCGTGCGTTCGCCAGCGCTACGACAGCGGCTGAAGCCGGCGGAGCG	264		
QY	571	tgattttacatgtcaacacttcgtaabtggaagtgaactgtcaaaaaataattct	630		
Db	265	GAAGCGCTACATGTCCGAGATGAAGCCCGCGCTCCGAGGCTTAAGAAACCTCCTTCT	324		
QY	631	caettgctcgaataatggacgcctataacttgcctacttacttacttcttccctcttg	690		
Db	325	CAACACAGCA-----CGTCACCTTCGTCGTCTACTCCACCTCTTTCATG	372		
QY	691	ggcagcagaggtagcagctgaagttcacatccctctgtctcttcttcttgagtcacacgc	750		
Db	373	GGCGCGGAGTGGCGGTTGTGCCAGCTCCGACGCGCCCTCTCTGGGTGCGAGCCGC	432		
QY	751	aacataatggacataattacttcaacttccatgatgaataagctatggctaatga	810		
Db	433	CACCGTGTGTGCATATACCACTTCTACTTCAACGGCTACGACGAGATCGACCGCG	492		
QY	811	atccaatgatccaaattggtccatcaacttcccgggttccactactggaaactcaga	870		
Db	493	TTCCAATGAA-----ATTGAGCTCCCTCGCTCCATCCCTGGAGCAGCGAG	540		
QY	871	tcttctcatttttacttctctatggtgcacaaagggagtcttcagttgcacttccacc	930		
Db	541	TCCTCCGAGCTTCTGCTGCC-TGCCAGCGCGGAGAGATTCGCGTTGATGA-----	590		
QY	931	attcaagaattatagacacattagatgctgaaccactcctagattcttgtaatac	990		
Db	591	--TGAAGGAGAAGCTGGAACCTTTAGACGGTGAAGAGAAGGCGAAGATTGGTGAACAC	648		
QY	991	attgatgaattagagcctgagccactcaatgcaattgaaggttataagttttatggaat	1050		
Db	649	GTTCGATCGTTGGAGCCCGATGCACTGACGGCTATTGATAGGTAGTTGATCGGGAT	708		
QY	1051	tggaccgttgattcctctgcttcttcttgggtggaatgacccttttagatgcttattgg	1110		
Db	709	CGGGCGCTGATTCCCTCCGCTTCTTGGACGGCAAGATCCCTCCGAACGCTCTACGG	768		
QY	1111	tgggtatcttttcaaaattcaaatgactata-----tgggaatggttaaaactcaaaagc	1164		
Db	769	CGCGCATCTTTTCGAAAATCGGAGGAGATAACTCGCTGGAGTGGTTGAACTCGAAGCC	828		
QY	1165	aaatcaatcagttgtttatataatttttggagctcctaataatcactatttagccaaat	1224		
Db	829	GAATCTTTCGTTGTTGATGTGCTGTTTGGAGCGCTTTTGGAGTTTTCGAAAGGCAAAAT	888		
QY	1225	ggagagatatcaaaaggttgatagacataggaagccgtttttatgggtgataaaga	1284		
Db	889	GGAAGAGATTGGGAAGGGCTATTACCTCGCGAAGGCCCTTTTATGGATGATACGAGA	948		
QY	1285	aaatgaanaagc-----aaagaagaagaataaaaagccttggttgattgaaga	1335		
Db	949	ACAGAAGAAATGACGACCGGCAAGGAAGAAGAAAGAGTAGTTGAGTTGCATTTGGGGA	1008		
QY	1336	attggaanaaattaggaanaattagttccatggtttcacaaacttggaattctcaaaacatcc	1395		
Db					
Db	1009	ATTGAAAAAATGGGAAAAATAGTGTGCTGGTCTCGCAGTTGGAGGTTCTGGCGCACCC	1068		
QY	1396	atcttaggatgttttcttctcatgttgatggaattcagccttagagagtttagcttg	1455		
Db	1069	TGCGTTGGGATGTTTCGTGACGCAATTTGGTGGAACTCGGCTGGAGAGCTTGAGTTG	1128		
QY	1456	tggagtgccagtttgccattcttccatggacagacatcaaatgacacaaatgccaacagt	1515		
Db	1129	CGGGATTCCGGTGTGGCGGTGCGCAGTGTGTTGATCAGACGAAATGCGAAGCTGAT	1198		
QY	1516	tgaagatgtgtggaagagtggaatgaagtgagaataataatgaagatggtgttgtaaaag	1575		
Db	1189	TCAGGATCGTGGCGGACAGGGTGAGAGTGAGATGAAGTAATGAAGGGGTGGGTTGATGG	1248		
QY	1576	tgaagaatacaaaaggtgtattgaattggttaagtgagagagagagaaggggaagaatt	1635		
Db	1249	ATGTGAGATAGAAGAGTGTGTGAGATGTTGATGGTGGGTGACAGACCAAACTAGT	1308		
QY	1636	gagaagaatctgaagaaatggaaagaattggctagagaagctgtgaaggaggtggatc	1695		
Db	1309	GAGAGAAAATGCCATCAATGGAAGACTTTGGCCAGACCAAGCCATGGATAGATGGATC	1368		
QY	1696	ttcacacaagaattaaaggcttttattgatgatgtgc	1734		
Db	1369	TTCACTCAACAATCTCAACGCCCTTCTCTCAAGTTGC	1407		
RESULT	5				
AB013598					
LOCUS	AB013598	2043 bp	mRNA	PLN	20-MAR-1999
DEFINITION	Verbena hybrida HGT8 mRNA for UDP-glucose:anthocyanin				
	5-O-glucosyltransferase, complete cds.				
ACCESSION	AB013598				
VERSION	AB013598.1	GI:4115562			
KEYWORDS	HGT8; UDP-glucose:anthocyanin 5-O-glucosyltransferase.				
SOURCE	Verbena hybrida petal cDNA to mRNA.				
ORGANISM	Verbena x hybrida				
REFERENCE	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Asteridae; Gentianaceae; Lamiales; Verbenaceae; Verbena.				
AUTHORS	1 (sites) Yamazaki,M., Gong,Z., Fukuchi-Mizutani,M., Fukui,Y., Tanaka,Y., Kusumi,T. and Saito,K.				
TITLE	Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin				
JOURNAL	J. Biol. Chem. 274 (11), 7405-7411 (1999)				
MEOLINE	99167509				
REFERENCE	2 (bases 1 to 2043)				
AUTHORS	Yamazaki,M., Saito,K. and Gong,Z.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-MAY-1998) to the DDBJ/EMBL/GenBank databases. Mami Yamazaki, Fac. Pharm. Sci., Chiba Univ., Lab. Mol. Biol. & Biotech.; Yayoi-cho 1-33, Inage-ku, Chiba, Chiba 263-8522, Japan (E-mail:mamiy@chiba-u.ac.jp, Tel:+81-43-290-2905, Fax:+81-43-290-2905)				
FEATURES	Location/Qualifiers				
source	1..2043				
	/organism="Verbena x hybrida"				
	/db_xref="taxon:76714"				
gene	26..1411				
	/tissue_type="petal"				
CDS	26..1411				
	/gene="HGT8"				
	/gene="HGT8"				
	/codon_start=1				
	/product="UDP-glucose:anthocyanin 5-O-glucosyltransferase"				
	/protein_id="BA036423.1"				
	/db_xref="GI:4115563"				
	/translation="MSRAHVVLLATPPAQGHINPQLQFAKRLANADIQVTFSTVYAMR				
	RMSRTAAGSNGLNIFVSFDGDDGKNYMSKSRGKALSDTLAANNVDQ				
	KSKRTFVYVYSHLFAAAKAVAREPHRLSALLMIPATVLDIFYFYFNGYSDIDAGSD				
	AHLPGSLPVLAQRLDPLSFLPSTHERFRSLMKLETLLEGEEKPKVLVNSFDALPDL				


```

CDS
/genes="AT4g14060"
join(4845..5031,5330..5598)
/genes="AT4g14060"
/notes="strong similarity to major latex protein type 1,
Arabidopsis thaliana; contains EST gb:A1994471.1, T43693,
A1099764, H36765, T43905, A1998305.1, R29829, T22825,
R90142, T41926, T22170, W43041, R65053, AA712594,
AA712597"
/codon_start=1
/product="major latex protein like"
/protein_id="CAB78448.1"
/db_xref="GI:7268111"
/translation="MATSQTYVTEVPLKGSARKHYKWRNENHLFPAIGHHIGQVTV
HDGEMSHGALKIMNYTLDGKPEMFKERKEDDENMAVTFVGLGHVMEOLKVYDITL
QFTOKSPDDIVCKITMTWEKRADDSPEPSNMYKLKLSLAADMDAHLKA"
4845..5031
/genes="AT4g14060"
/number=1
5032..5329
/genes="AT4g14060"
/number=1
5330..5598
/genes="AT4g14060"
/number=2
complement(9562..9654)
/number=1
/genes="AT4g14070"
complement(join(9562..9654,9750..9807,9883..9998,
10097..10176,10276..10435,10594..10731,11065..11145,
11247..11300,11383..11541,11653..11758,11882..11949,
12156..12305,12439..12554,12626..12834,13288..13360,
13558..14072,14674..14859,15291..15425))
/notes="strong similarity to AMP-binding protein - Brassica
napus; Contains putative AMP-binding domain signature
AA255-266; contains EST gb:A1999282.1"
/codon_start=1
/product="AMP-binding protein"
/protein_id="CAB78449.1"
/db_xref="GI:7268112"
/translation="MASTSLGPTLLSYGSPSRQFPDFGFRLLISGHESVRIPSRFF
REKVEVSPFFLESSFSGDAALRSSEKAYVDIHRSAKYGDYVALVDYDHPFLK
LTYKLEQELIPFAGELVYKADKALFADNSRWLSDQGANVAVRGSRSSVEE
LLQIYRHSEVAIVDNPFFNRIAESFTSKASLRFLLIWMGKSSLVTGOMGIPYIS
HAIINOGESEKAKLSNDTSRNQFIDSDDTAAIMYTSGTGNPKGYMLTNRLL
HQIKLSKYPQAGDKFLSMLPSWHAYERASEYFIETCGVEQMYTSIRYKDLKRLY
QPNYIVSVPLVETLYSGIQKQISASSAGRKFLALTIKVSMAVMEMKRIYEGMCILTK
EQKPPYIVAFVDLWLVIAALWPLHMLAKLIYKKIHSSIGISKAGISGGSLPI
HVDKFEAIGYILQNCYGLTETSPVVCARTLSCNVLGASGHPMHGTEFKIVDPTNNV
LPKSGKILIKVRGPOVMGYKNPSTTKQVLESQWNTGDTGWIAPHHSKSGHCC
GVLYLGRKADIVLSTGENVEPLEBAAMRSRVIEQIVVIGQDRRLRGLIIPNKE
EAKVDFPESKELASLVLTQELRKWTSECSQVGPVLVDDPFTIDNGLMTPMTKIRR
DMVVAYKKEEDQLYS"
complement(9655..9749)
/number=1
complement(9750..9807)
/number=2
complement(9808..9882)
/number=2
complement(9883..9998)
/number=3
complement(9999..10096)
/number=3
complement(10097..10176)
/number=4
complement(10177..10275)
/number=4
complement(10276..10435)
/number=5
complement(10436..10593)
/number=5
complement(10594..10731)
/number=6

intron
complement(10732..11064)
/number=6
complement(11065..11145)
/number=7
complement(11146..11246)
/number=7
complement(11247..11300)
/number=8
complement(11301..11382)
/number=8
complement(11383..11541)
/number=9
complement(11542..11652)
/number=9
complement(11653..11758)
/number=10
complement(11759..11881)
/number=10
complement(11882..11949)
/number=11
complement(11950..12155)
/number=11
complement(12156..12305)
/number=12
complement(12306..12438)
/number=12
complement(12439..12554)
/number=13
complement(12555..12625)
/number=13
complement(12626..12834)
/number=14
complement(12835..13287)
/number=14
complement(13288..13360)
/number=15
complement(13361..13957)
/number=15
complement(13958..14072)
/number=16
complement(14073..14673)
/number=16
complement(14674..14859)
/number=17
complement(14860..15290)
/number=17
complement(15291..15425)
/number=18
complement(16137..16174)
/number=1
16137..17732
/genes="AT4g14080"
complement(join(16137..16174,16252..17604,17687..17732))
/notes="the cDNA A26668 is cognate to this gene; gene A6
for anther-specific protein - beta-1, 3-glucanase homolog;
Contains Glycosyl hydrolases family 17 signature
AA274-287; Prokaryotic membrane lipid protein lipid
attachment site AA8-18; contains EST gb:AW004339.1,
Z18400"
/codon_start=1
/product="A6 anther-specific protein"
/protein_id="CAB78450.1"
/db_xref="GI:7268113"
/translation="MSLLAFFLFTLVFSSSCSATRFQGHRYMORKTMDLASKIGI
NYGRGNLPSPYOSINFIKSIKAGHKLYADPESLTLLSQTNLYVTIVPNHQITA
LSSNQTIADQVRYTNILPYYPQQTQIRFVLVGNELLSYNSGVNLYPARKKIVNSLR
LHGHNKIKVGTPLAMDLSRSPSPNSGTFREITGVPMLPLLAFLNGTNSFTFLNVHP
YFRWSRNPMTSLDFALFQGHSTYTDPTQGLVYRNLLDQMLDSVLFAMTKLYPHMRL
A1SETGMPTGDIETGAILHFDGSPYIDVDFTPTGTGNPLKPTNNVYKGVQWCVF
NENKSGSGTQRHWGILHFDGSPYIDVDFTPTGTGNPLKPTNNVYKGVQWCVF
VEGANQTELEETLRMACASNTTCAALAGRECYEVSIVYHSAVYALNSYWAOFRIQS
IQCFNGLAHETTTPGNDRCCKFPSTLT"
complement(16175..16251)

```

```

exon      /number=1
complement(16252..17604)
Intron    /number=2
complement(17605..17686)
exon      /number=2
complement(17687..17732)
exon      /number=3
18189..18494
/gene="AT4g14090"
/number=1
18189..25761
/gene="AT4g14090"
Join(18189..18494,18694..19018,19571..19724,
19772..19833,21489..21524,22014..22792,22844..22948,
23060..23200,23280..23566,24375..24631,24686..25140,
25192..25311,25445..25614,25665..25761)
/gene="AT4g14090"
/number=1
note="similarity to proton-coupled peptide transporter
PEPT, Rattus norvegicus; Contains prokaryotic membrane
lipoprotein lipid attachment site AA535-545;Prokaryotic
membrane lipoprotein lipid attachment site AA955-965;
contains EST gb:AI998193.1"

```

```

Query Match      14.48; Score 303.2; DB 8; Length 198944;
Best Local Similarity 56.28; Pred. No. 7e-51;
Matches 796; Conservative 0; Mismatches 543; Indels 77; Gaps 9;

QY 350 cctcatgttcatttaacacattccagcaagggccatatttaaccagcacttcaattt 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21211 CCACATTACTTCTGTTAACTATCCACGCAAGGTACATCAACCCGCGCTTCAACTA 21152

QY 410 gccagaatcttgcagaatggcatagaagtgcacatttttctacaagaattatgccaa 469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21151 GCAACCGCTCATCCACACGGTGCACCGTACATCTCCACCGCATCTCTGCTCAC 21092

QY 470 agccgtatggatgaaaaatccattcttaatgcacaaaaggattgaatttcattccatt 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21091 CGACGTATGGGCGAGCCACC-----TTCCACAAAAGGTCTATCCTTCGCTTGGTTC 21041

QY 530 tccgatggttttgatgaaggttttgatcattcaaaagaccctgtattttacatgtccaaa 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21040 ACCGATGATTCGACGCGTCTCAGTCATTCGACAGCCAGAAAATCTACATGTCGNA 20981

QY 590 cttcgttaaatggaagtgaactgtcctacttactccttcttcttcttcttcttcttct 640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20980 CTCAAACGATGTGGTTTAAACGCGCTCGAGAGACATCATCAAAAGCAATCTTGACGCCAC 20921

QY 641 gaaatggacagcctataacttgcctacttacttacttcttcttcttcttcttcttcttct 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20920 ACCGAAACAGACCTATACCGGGGTAACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 20861

QY 701 gtacacgtgaagttcacatccctcttcttcttcttcttcttcttcttcttcttcttcttct 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20860 GTAGCGCGTGAAGTTTACCTCCCACTACACTTCTCTGATTTGAACACGACTACTGTACTA 20801

QY 761 gacatatattctcaacttcttgatataaaaagcctatgcttaaatccaatgat 820
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20800 GACATCTACTACTACTCTCAACCTCTTACAAACATCTCTTCGACGT----- 20749

QY 821 ccaaatgggtccattcaacttccgggcttccactactggaactcagatcttcttca 880
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20750 -----TGAACCGATTAAATTAACGAAATGCCACTGTATCACCGGTGACCTCCCGTCC 20696

QY 881 ttttacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 940
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20695 TTCTTCAACCTTCGAGGCAATACCGGTGACGTCTGTGACTCTA-----AGAGAA 20645

QY 941 ttgatagacacattagatgctggaaccactcctaaagtatttttgatatacatttgatgaa 1000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20644 CATATCGAAGCTCTCGAAACGGAATCAAAACCTTAAGATCTTCTTTAAACATCTCTGCT 20585

QY 1001 tttagacctgagcactcaatggaagttaaatgttttatggaattgaccgcttg 1060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 20584 TTGGAAACAGCATGCTTTAACTCTGTTGAGAAACTCAAGATGATCCCAATCGACCGTTG 20525
QY 1061 attcctctgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20524 GT---TTCCTCTCCGAGGTAACACCGATCTTTTCAAAATCTTCCGACG----- 20477
QY 1121 ttccaataatcaaatgactatataatggaatggtaaatcaaatcaaatcaaatcaaatg 1180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20478 -----AGGATTACACGAATGTTAGACTCGAAGCTCGAGACTCAGTGATT 20432
QY 1181 tatatatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20431 TACATTTCTCTAGGACACACCGCGATGATTACAGAGAAACACATGGAAGCGCTTACT 20372
QY 1238 aaagggttgatagacataggaagccgcttttttcttcttcttcttcttcttcttcttctt 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20371 CACGGCGTCTTACTTACAAACAGACCGCTTTTATGATCGTGAGGA---GAAAAATCC 20316
QY 1298 aaagaagaagaataaaagaactgtgtgttatgtatgaagaattgga-aaaaataggaataa 1356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20315 AGAAGAGAAGAAGAAGATCGGTTTCTTGAATGATCAGAGGAAGTATCGAGGATTGGT 20256
QY 1357 agttccatgtgttcacaaactgaactgaacttcaaaacatccatctttagagattttgttc 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20255 GGTGGGATGCTTCTCAGACAGCTGTTTGGCGCATGTGCTGTGGGATGTTTGTGAC 20196
QY 1417 tcaattgtgatggaattcagccttagagagtttagcttcttcttcttcttcttcttcttctt 1476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20195 TCATTCTGCTTGAATTCGACGTTGGAGAGTTTACAGAGTGGTCTCCGCTGCTGCTT 20136
QY 1477 tctcaatggacagacataatgcaaatgcaaatgcaaatgcaaatgcaaatgcaaatgca 1536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20135 TCCGCGATTTGCTGTATGATGATGATACACGCGAAGCTTGTGAGGATACGTGGAGGATTGG 20076
QY 1537 agtaagtgagaataaataagaatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtat 1596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20075 AGTGAAGGTGAAGGTGGGAGGAGAGATCTGATCGGAGGAGATTTAGAAGGTGTTT 20016
QY 1597 tgaattgttaattggtgagagagagagagagagagagagagagagagagagagagagag 1656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20015 GGAAGAGGTGATGAGTGGTGGAGAGAGCGGAGAGATGAGAGAGATTCAGAGAGATG 19956
QY 1657 gaaagaattggatagagaagctgtgagaagagagagagagagagagagagagagagagag 1716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19955 GAAGCGGATGGCTGTGTGATCGGCGACGGAAGGTGGACCGCTCGGATTTGAATCTTAAAG 19896
QY 1717 ttatttgatgattgtccaaagggttttaattt 1752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19895 TTTTGTGGACGAGGATGAGTAGTGGGTTTAAAGTTT 19860

RESULT 7
ATFCAO/LOCUS 200576 bp DNA PLN 28-JUN-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
ACCESSION 297335
VERSION 297335.2 GI:5280985
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicotyledons; Rosidae; eucosids II; Brassicales; Brassicaceae;
Arabidopsi.
REFERENCE 1 (bases 1 to 200576)
AUTHORS Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terry,N.,
Kreiss,M., Kavanagh,T., Entian,K.D., Rieger,M., James,M.,
Puidomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
Jones,J., Palme,K., Ansong,W., Delseny,M., Bancroft,I.,
Mewes,H.W., Schueller,C. and Chalwatzis,N.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 200576)

```


AUTHORS EU Arabidopsis sequencing, project.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuellemips.biochem.mpg.de,mayeremips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

COMMENT
 On Jun 30, 1999 this sequence of analysis replaced gi:2244747.
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with AF041 at the 3' end.

FEATURES
 source
 1..200576
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 1..63289
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="BAC TAMU20K8"
 1753..4428
 /gene="dl3000w"
 join(1753..3194,3261..4428)
 /gene="dl3000w"
 /note="similarity to Cf-2.2, Solanum pimpinellifolium"
 /product="start=1"
 /protein_id="CAB10171.1"
 /db_xref="GI:2244748"
 /db_xref="SPTREMBL:O23253"
 /translation="MTIINSLCLIFCLNSILVIAKDCLPQDRDALLFKNFESIP
 SPDSLMILQITAKRANTCCSMGISCDFKTVGVVLDLGNLSNDLGRNSNSIF
 RQHLQSLDSLNDUSCTLPDSSNGFYLRVNLGLNGLFGEIPTSLRSLSYLTDL
 SYNDLDTLSMGLKHLRLSLTSCKFTGKIPSSLGLNLYLTLDLSLWNYFTGL
 PDSMGLKRLVNLHRCNFRFKIPTSLGSLNLTDLISKEFTSEGPDSMSLRFL
 TQMLMLLSLTVNLDSSNQFKAMLPSSMSLSKLEAFDISGFSFGTIPSSFLML
 PSILKLDGTCNFGPLKIGNISPSNLOELYIGENNINGIPRSILKVLGALSLS
 FWTGQIVDFSTFLQKLSRLDLSGINLTSSSHHPLSHMHMLILSSCNISQPKFL
 ENQTSILYDIANSIOLEGQPEMLWRLPTLSFIASDNKFSCEIPRAVCEIGTVLSNN
 NFSGSIPTCFEISNKTILHLRNLSLGVIPESLGLYLRSLDVSNRSLSGQFPKSL
 INCSYQFNVEENENINDTFPSLKLNLQNLVLRNNEEHPGFSPGDSLSFKRFL
 FDIENRFGVLPSTDFVGMVMSFVDIDINTPGFTVGDQDESFKHSVYLTIKGLN
 MELVSGFITYITDVSNLRLEGPDISIGILKELIVLNMNNAFTGHIPPLSNLSN
 LQSLDLSQRLSGSIPGLGELTFLARNFYNMLEGPIPGQTIOQNSSSFAENPG
 LCGAPLQKCGGEEDKEKEKDKGLSWAAAIQYVPGLCGLAIGHILTYSYKRDWF
 MRFSCFSSPLAKK"
 1753..3194
 /gene="dl3000w"
 /number=1
 3195..3260
 /gene="dl3000w"
 /number=1
 3261..4428
 /gene="dl3000w"
 /number=2
 complement(5905..6149)
 /number=1
 5905..7913
 /gene="dl3005c"
 complement(join(5905..6149,6243..6344,6436..6826,
 7236..7913))
 /note="strong similarity to glycine
 hydroxymethyltransferase (EC 2.1.2.1); contains EST
 gb:1565775"
 /codon_start=1
 /product="hydroxymethyltransferase"
 /protein_id="CAB10172.1"
 /db_xref="GI:2244749"
 /db_xref="SPTREMBL:O23254"
 /translation="MEPVSSWGNSTLSVSDPEIHDLIEKKEKRCQRCIELIASENFTS
 FAVIEALGSALTNNKYSBEGIPGNRYGGNEFTIDEIENICRSRALEAFHCDPAANGVNVQ
 PYSGPSANFAAYTALLQPHDRIMGLDLPSSGGLTHGYTSGGKISATSYFESLPYK
 VNETTGYIDYDLKEKALDRPRLKLLICGGSAYPRDWYFARFRAIADKVGALLCDMAH
 ISGLVAQAANPFYCDVVTTHKSLURGPAGNIFYRKGPKPYPKKGQEGAVYDFE
 DKLNFAAFALOGQPHIQHIGALVALKAQANTPGFYVAKQVKANAVALGNLMSKGY
 QVTNGTEHNLVMDLRPLRLGLTKGNKVKELCDLCSITLKNKAVFCDSSALAPGSGVICA
 PANTSGLVEXKDFEQIGFELSRANVTLLDIOKTYGKLLKDFNKGVLNKKDLQDLKADY
 EKFSASYEMPGFLMSEMKYK"
 complement(6150..6242)
 /number=1
 complement(6243..6344)
 /number=2
 complement(6345..6435)
 /number=2
 complement(6436..6826)
 /number=3
 complement(6827..7235)
 /number=3
 complement(7236..7913)
 /number=4
 12823..14568
 /gene="dl3010w"
 join(12823..13533,13822..14568)
 /gene="dl3010w"
 /note="strong similarity to adenosylhomocysteinaseE (EC
 3.1.1.1)"
 /codon_start=1
 /product="adenosylhomocysteinase"
 /protein_id="CAB10173.1"
 /db_xref="GI:2244750"
 /db_xref="SWISS-PROT:O23255"
 /translation="NALVEKTSRGREYKVKDMSQDQGRLELEAEVMPGLMACRT
 EFQSPFPFGARTGSLHMTQTAVLITETALGAENRWKSCNFTQDHAHAARIAD
 SAAYFAWKGETQEWCTERALDMPGGPDLIVDGGDNATLLIHGVKAEEIFERT
 QVDPDTSTNPEFQIVLSIIEKGLQVDPKRYHKMKRLVGVSEETTTGVRKLYHQOQ
 NGTLLPAINVDSVTKSKFNLGYCRSLPDGLMRDVMIAKVAVICGYDGVGKG
 CAAMKTAGARYIVTEIDPICALQALMEGLQVLTLEDVYSEADIVFTVTGNKDIINVD
 HMKMKNAIVNIGHFNDIMLGLETYPGVKRTITKPTQDRWVFPETKAGIIVLAE
 GRMLNKGATGHPSPFVMSCFSTNOVIAOLELWNEKASKYEKKVYVLPKHLDEKVAL
 HLKLCARLTGKLSKQSDVYSPIEGPKYPHYR"
 12823..13533
 /gene="dl3010w"
 /number=1
 13534..13821
 /gene="dl3010w"
 /number=1
 13822..14568
 /gene="dl3010w"
 /number=2
 complement(16093..16270)
 /number=1
 16093..16501
 /gene="dl3015c"
 complement(join(16093..16270,16317..16501))
 /note="weak similarity to Walleye dermal sarcoma virus gag
 polyprotein"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB10174.1"
 /db_xref="GI:2244751"
 /db_xref="SPTREMBL:O23256"
 /translation="MFNSTALVIFAIFLLISADAFPTSPNGSIDAMLRINSIIGED
 EDLMPTEIGRRVLMKAKRNOELLTVALDKLILTAEAVTPLDPALETQTTSTHEGKI
 QETAKIGSFIAEKSLYD"
 complement(16271..16316)
 /number=1
 complement(16317..16501)
 /number=2
 complement(join(18573..18914,18997..19137,19212..19534,
 19603..19719,19762..20123))
 /note="weak similarity to fusion protein la/1b - sugar
 beet"
 /codon_start=1

REFERENCE	Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1. (bases 1 to 195068)				
TITLE	EU Arabidopsis sequencing, project.				
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk				
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV41 at the 5' end and an overlap with ATCHRIV43 at the 3' end.				
FEATURES	Location/Qualifiers				
Source	1..195068				
	/organism="Arabidopsis thaliana"				
	/variety="Columbia"				
	/db_xref="taxon:3702"				
	/chromosome="4"				
exon	9917..10065				/number=1
	/gene="At4g15560"				
CDS	join(9917..10066,10350..10475,10559..10874,10981..11252,11340..11613,11691..11998,12081..12362,12441..12725,12813..12953)				/number=1
	/note="GT/AC sites ignored to preserve continuity; light inducible gene required for chloroplast development; Contains prokaryotic membrane lipoprotein lipid attachment site AA452-462; Prokaryotic membrane lipoprotein lipid attachment site AA591-601; Transketolase signatures AA102-121; Transketolase signatures AA505-521; contains EST gb:AW004332.1, AI997953.1, W43562, N96015, W43388"				
	/codon_start=1				
	/product="DEF (CIA1) protein"				
	/protein_id="Cab78598.1"				
	/db_xref="GI:7268304"				
	/translation="MASSAFAPSVIITKGLSTDCSTSLSSRSLSVTDLPSPCLK PNNSHNSRAKVCASLAEGKGYNSRPPTPILLDTINYPHMKNLVSKELKQLSDSLR SDVIFVSKTGKGLSGSLGVVELTVALYIFNTPQDKILMDVGHQSTPHKILVTRGRK MPTMQNGLSGFTKRGESEHDFCTGHSSTTISAGLGMVGRDLKGNKNNVAVITGD MPALRELVKGMTRQGGPMHQAARVDEYARGMTSGTSSLFELGIYYICPDVG HNIDDLVALIKVEKSTRITGPVLIVHVTKEGRGYPVAERADDKYGVVKFPDPTGRQF KTTNKQSYTYFAALVAEAEVDKDVVAIAAAMGGTGLNLFORPFRPCPDVGIAE QHATVFAAGLACGLKPCFCAIYSSPMORAYDQVVDVLDLKLVPFRFAMDAGLADG PTHCGAFDVFNACLPNMTWAPSDADLFNNVATAVIDRSCFRYPGNGIGVAL PFGKGVPIEIGKRLKEGERVALLGSAVQSLGAAVHBERGLNVTYVADARCK PLDRALIRLSAKSHVELLITVEEGSTGGFGSHVQVFLADLGLLDGLKLRWPRPDRYI DHGAPADQLAEAGLMPSHTAATNALNLIGAPREALF"				
gene	9917..12953				
	/gene="At4g15560"				
intron	10067..10349				
	/gene="At4g15560"				
exon	10350..10475				/number=1
	/gene="At4g15560"				
intron	10476..10558				/number=2
	/gene="At4g15560"				
exon	10559..10874				/number=2
	/gene="At4g15560"				
intron	10875..10980				/number=3
	/gene="At4g15560"				
exon	10981..11252				/number=3
	/gene="At4g15560"				
	/number=4				
	/gene="At4g15560"				
	/number=4				
	11340..11613				
	/gene="At4g15560"				
	/number=5				
	11614..11690				
	/gene="At4g15560"				
	/number=5				
	11691..11998				
	/gene="At4g15560"				
	/number=6				
	11999..12080				
	/gene="At4g15560"				
	/number=6				
	12081..12362				
	/gene="At4g15560"				
	/number=7				
	12363..12440				
	/gene="At4g15560"				
	/number=7				
	12441..12725				
	/gene="At4g15560"				
	/number=8				
	12726..12812				
	/gene="At4g15560"				
	/number=8				
	12813..12953				
	/gene="At4g15560"				
	/number=9				
	14413..14436				
	20338..24557				
	/gene="At4g15570"				
	20338..20373				
	/gene="At4g15570"				
	/number=1				
CDS	join(<20338..20373,20767..20852,20962..21012,21256..21367,21467..21572,21653..21750,21989..22117,22182..22274,22625..22678,22957..23015,23122..23206,23286..23765,24283..24269,24366..24557)				
	/note="similarity to SEN1 protein, yeast; Contains ATP/GTP-binding site motif A (P-loop) AAL3-20; contains EST gb:N65040"				
	/codon_start=1				
	/product="SEN1 like protein"				
	/protein_id="Cab78599.1"				
	/db_xref="GI:7268305"				
	/translation="VGLSRKSFVLQGGPGTGKTQTILSLIGATMHATPARVOSKCTD HEVKGIOHTIOEKYHNHGRASFWILGVNPRDAIMPEDGDDFFPTSGNELKEPVNA SRKTRULRVLCAPSNSALDEIVLRLLSLGXDNRATYTPKIVRIGUKAHHSVASVSL DHLVAQRSAIDKPKQGTGTGDDISIRTAILEAAIIVGFNFRINVFATLSFGSAL LAKSNRQFVVIIDEAAQAVEPATLIPLATRCKQVFLVGPDKLPATVIYSTVAQDSGY GTSMERFLQAGYPVKMLATQYRMHPELIPISQTKPKIVLSLILTPMIASTYFTLOTIF LTENQIRSPSKQFYEGALDGDSDIEATQTRDMHKYRCFGFCFFDHEGKSHQPCA TGSRVNLDEVEFVLLIYHRLVTMYPELAKSSQLALISPYNQVTKFKDRFKEMGTEA EKYVDINTVDGQVVGSAATLAKSDPLMKNLIESAEQRNLFKVKSKPLNFFSEENLET MKLTEDHEIPDAPLIEDESLVPVPIGGDDDFGDDGDDQDDVVMAGED"				
	20374..20766				
	/gene="At4g15570"				
	/number=1				
	20767..20852				
	/gene="At4g15570"				
	/number=2				
	20853..20961				
	/gene="At4g15570"				
	/number=2				
	20962..21012				
	/gene="At4g15570"				
	/number=3				
	21013..21255				
	/gene="At4g15570"				

```

exon          /number=3
21256..21367
/ gene="AT4g15570"
/ number=4
21368..21466
/ gene="AT4g15570"
/ number=4
21467..21572
/ gene="AT4g15570"
/ number=5
21573..21652
/ gene="AT4g15570"
/ number=5
21653..21750
/ gene="AT4g15570"
/ number=6
21751..21988
/ gene="AT4g15570"
/ number=6
21989..22117
/ gene="AT4g15570"
/ number=7
22118..22181
/ gene="AT4g15570"
/ number=7
22182..22274
/ gene="AT4g15570"
/ number=8
22275..22624
/ gene="AT4g15570"
/ number=8
22625..22678
/ gene="AT4g15570"
/ number=9
22679..22956
/ gene="AT4g15570"
/ number=9
22957..23015
/ gene="AT4g15570"
/ number=10
23016..23121
/ gene="AT4g15570"
/ number=10
23122..23206
/ gene="AT4g15570"
/ number=11

Query Match      12.5%  Score 263;  DB 8;  Length 195068;
Best Local Similarity 56.0%;  Pred. No. 7.2e-43;
Matches 674;  Conservative 0;  Mismatches 445;  Indels 84;  Gaps 6;

Qy 556 tcattcaaaagaccctgtattttacatgcacaacttcgtataatgttggaagtgaactgt 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4734 TCGTCAAGACGCCACTGGAAACTTCATCTGAGATGAGACGCGTGGCAAGAGACACT 4675

Qy 616 caaaaaataattctcaactgtctgaaaaatggacagcctaataacttgcctacttacc 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4674 AACCGAACTAATGAAGATAAACCGGAACAAACAAACAGGCCCTTTTACTTGGCGTTTACAC 4615

Qy 676 catttctctcttggcagcagaggtgagcagctgaagttcacatccctctctctctct 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4614 GATTCTCTCACTTGGGTGGCTAGCGTAGCGGTGAGTTTCATCTTCTCTCTCTCTCTCT 4555

Qy 736 ttgagtcacacagcaacaatttgacatatattactcaacttccatgatataaaa 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4554 TTGGGTCCAAACCAAGTAAACAGTCTTCTCCATTTTTCACATTTTACATTTCAATGGCTACGAAGA 4495

Qy 796 agctatggctaatgaatccaatgatcccaattggtccattcaactcccccgggtccact 855
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4494 TGCATCTCAGAGATGGCTAATACCCCTCTAGTCTATTAATTAACCTTCTCTGCACT 4435

Qy 856 actggaaactcagatcttctctctctctctctctctctctctctctctctctctctct 915
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 4434 GCTTACTGTCCGTGATATTCCCTTCTTTCATGTCTCTTCCCAATGCTACGGTTTCTT-- 4375
Qy 916 agttgcacttccacccattcaagaattgatagacacattagatgctgaaccactctaa 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4376 -----CTACCGCGTGTTCGAGAACAGATTGATTACTAGAGGAAGAAATAAACCCCTAA 4324
Qy 976 gattctgtgaatacatttgatgaattagagcctgagcactcaatgaattgaagatta 1035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4323 GATCCCTCATCAACACTTTCGAAGAGCTTGAGCCAGAACCCATGAGCTCGGTTCCAGATAA 4264
Qy 1036 taagttttatggaaattggaccgttgattccctctctctctctctctctctctctctct 1095
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4263 -----TTTCAAGATTGTCCCTGTGCGTCCGTTACTAACGTTGAGAACGGA----- 4217
Qy 1096 agatgcttccatttgggtgatctttttcaaaaattcaaatgaactatgaatgaagttaaa 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4218 -----TTTTTCGAGTCGCGGTGAATACATAGAGTGGTGGGA 4183
Qy 1156 ctcaagcccaattcatcagttgtttatatatatcttttgggagttctaataatccatctat 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4182 TACTAAAGCGGATTCGTCTGTCTTATGTTCGTTTCGGGACGCTTGCCTGTGTAGCAA 4123
Qy 1216 tagcacaatggagagatatacaaaaggttgatagacataggaagccgtttttatgggt 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4122 GAACAGCTTGTGGAGCTTGTAAAGCGTTGATACAAAGTCGGAGACCATTTCTTGTGGGT 4063
Qy 1276 gataaaagaaaatgaaaaggcacaagaagagagataaaaaagcttg----- 1322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4062 GATTACGGATAAGTCGTACAGAAATAAAGAAAGATGAGCAAGAGAAAGAAAGATTGCAT 4003
Qy 1323 -----gttgattgaaagattgaaaaaataggaaaaatagttccatgggttccacact 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4002 AAGTAGTTTCAGAGAAAGAGCTCGATGAGATAGGAATGGTGGTTTCATGGTGTGATCAGTT 3943
Qy 1378 tgaagttctaaaaacatccatctttaggatgttttttctctcattgtgtggaattcagc 1437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3942 TAGGGTTTGAATCATAGATCGATAGTTGTTTCGTGACGCAATTCGCGGTGGAACATC 3883
Qy 1438 cttagagagtttagcttggagtgccagttgctcattctcctcaatggacacatcaat 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3882 GCTGGAGAGCTTGGTTTCAGGAGTTCGGGTGGCGCTTCGCAANTGAATGATCAGAT 3823
Qy 1498 gacaaatgccaaacaagttgaaagtgtgtgaaaaagttggagtaagagtga----- 1547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3822 GATGACCGGAAGCTTTTAGAAGATTGTTGGAAAAACAGGTGTAAAGAGTGTAGGAGAGAA 3763
Qy 1548 --gaataaatgaagatggtgtgtgaaagtgaagatacaaaaggtgtattgattggt 1605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3762 GGAAGAAGAAGAGAGTGTGGTGTGATGAGAGATACGGCGGTGCATTTGAGGAAGT 3703
Qy 1606 aatgatggaggagagaaaagggggaagaattgagaagaatgctgaagaaatggaagaatt 1665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3702 TATGGA-----AGACAAGCGGAGAGAGTTTAGAGGAAATGCCACGAGTGGAAAGATT 3649
Qy 1666 ggtcagagagctgtgaagaaaggtggatcttcacagaatttaagagctttttatga 1725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3648 AGCGCGGAGCGCTGTGAGAGAGAGGAGCTCTCTCTTTTATCATCTCAAAAGCTTTTGTGCA 3589
Qy 1726 tga 1728
    |||
Db 3588 TGA 3586

```

```

RESULT 9
ATCHRIV41/c
LOCUS ATCHRIV41 197419 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 41.
ACCESSION AL161541
VERSION AL161541.2 GI:7268270
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```


Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE AUTHORS

1 (bases 1 to 205065)
Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N., Kreis, M., Kavanagh, T., Entlan, K.D., Rieger, M., James, R., Puigdomenech, P., Hatzipoulos, P., Obermaier, B., Duesterhoft, A., Jones, J., Palme, K., Ansong, W., Delseny, M., Bancroft, I., Mewes, H.W., Schueller, C. and Chalwatzis, N.

JOURNAL

REFERENCE

2 (bases 1 to 205065)
EU Arabidopsis sequencing, project.

AUTHORS

TITLE

JOURNAL

Submitted (25-Jun-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuellemips.biochem.mpg.de, mayer@mips.biochem.mpg.de
Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

COMMENT

On Jun 30, 1999 this sequence version replaced gi:2244901.
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATPCA3 at the 5' end and an overlap with ATPCA5 at the 3' end.

FEATURES

source

1. 205065
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1. 14029
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="cosmid YUP5B11.B1G15/YUP5B11.B1G19"
4257. 7108
/gene="dl3766w"
4257. 4662
/gene="dl3766w"
/number=1
join(<4257. 4662, 6078. 7108)
/gene="dl3766w"
/note="strong similarity to fatty acid hydroperoxide lyase
- Capsicum annuum; contains EST gb:949688"

gene

4257. 7108

exon

4257. 4662

CDS

join(<4257. 4662, 6078. 7108)

/product="hydroperoxide lyase (HPOL) like protein"
/protein_id="CAB45989.1"
/db_xref="GI:5281016"

/translation="NLSRAITPLTASPPYNAVGLAVGWTILGPRFGDPKFFRT
RASKYSTVFRNIPPTFFFGNVPNIYAVLDVKSFLHFDMDLDVDRDLVIGDRFP
SLGTFGVGRVYLDTPTEPKAKVTVYVALLGLAFAMETLKRSKYKWLRLNLNIF
WGTFSEIKNGAASYIPLOKICFIFCLASLGVADSPDAENGWKTINTWLAQ
VIPTAKLVVPOPLEEILHTWTPYSLIAGNYKKLYNFIDNAGDCLRLGDEFGILT
RDAIONLFLVGLFNAGYGFVFLPSLITGRITGNSGLQIRIETVRVRCGSGDLNF
KTVNEMELKVSIVYETLRFSPVPLQFARKKDFQISSHDAVFEVKKGLCYQPLV
MRDANVDEPEEFKPDYVYGTGSELLNLYWSNGPQTGTPSASNKQCAKADIVITLA
SLLVADLRLYDTITGDSGSIKAVYKAK"

4663. 6077
/gene="dl3766w"
/number=1
6078. 7108
/gene="dl3766w"
/number=2
8083. 8712
/gene="dl3766w"
/number=1
8083. 9807
/gene="dl3766w"
join(8083. 8712, 8790. 8899, 9225. 9395, 9521. 9574,
9654. 9807)
/gene="dl3766w"
/note="predicted"

intron

4663. 6077

exon

6078. 7108

exon

8083. 8712

gene

8083. 9807

CDS

join(8083. 8712, 8790. 8899, 9225. 9395, 9521. 9574,
9654. 9807)

/codon_start=1
/product="hypothetical protein"
/protein_id="CAB45990.1"
/db_xref="GI:5281017"
/translation="MSFRSKTSKTTQOPSSYQQQODAMYSYGVTVSDVLLQIHRC
RAHLINGSVAVELAAAGDFELVOVLDNSVALAMVVRIGNDLQWPVVKIDPSRDY
LFTLPVKDEPISYGIPTTFPIDENDIVFVNSIELDDDFLRKNSCFSSSSSSSSSS
SSSVNNGIDMKFAFKLEDYNNVAKAIAAGTGTHIIRGMFKCSNATVQVHKGGMIM
ITAEKNGKASSKRNATNNKQINKNLRVRKLSRTEKLSKTLNKGVGVSNGWGP
VYKSKGKAPFAMVPEVLLASLDLSALSKETNSFCYVOSSTNMVSRIGESAGEATKDV
LGTVGAAGTAWNVFNIRKAFSPSSVTSGLTKNASRK"
8713. 8789
/gene="dl3768w"
/number=1
8790. 8899
/gene="dl3768w"
/number=2
8900. 9224
/gene="dl3768w"
/number=2
9225. 9395
/gene="dl3768w"
/number=3
9396. 9520
/gene="dl3768w"
/number=3
9521. 9574
/gene="dl3768w"
/number=4
9575. 9653
/gene="dl3768w"
/number=4
9654. 9807
/gene="dl3768w"
/number=5
complement(10065. 10511)
/note="similarity to L.esculentum Tfm5 gene"
/codon_start=1
/product="glycine-rich protein like"
/protein_id="CAB10324.1"
/db_xref="GI:2244903"
/db_xref="SPTREMBL:O23398"
/translation="MKIKTKRAHFVYNYAAKLAVALLLITLLVNTTVVAKGGGGGAH
GGGVHVSVGGAAHSAVGGHSGGGHSGGGHSGGGHSGGGHSGGGHSGGGHSGGGH
EEGCHGIGRGGMVHRPATRNGSTLTPTTFVDGTTTFVMIIFLFN"
complement(10065. 10511)
/number=1
10065. 10511
/gene="dl3770c"
join(12306. 12597, 13233. 13495, 13554. 13835, 13944. 14000,
15045. 15290, 15396. 15802, 16095. 16280, 16380. 16453,
17015. 17346)
/gene="dl3775w"
/note="similar to hypothetical protein C02F5.7,
Caenorhabditis elegans; contains EST
gb:1217311.937779.315469"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB10325.1"
/db_xref="GI:2244904"
/db_xref="SPTREMBL:O23399"
/translation="MDKPYGVASVMSGIDRSAGKIDIDLEMVGCEATLYPGLSYGENO
LRMGFKRVYGLISAQLLLTTLISAVVLPNPVNDLTGSPGILLFLICIVPFTLLSFE
FVSTEYLDVFLHSRTSALYFMSYLAITREIAVIDWEFNFWIMPLHLYHKHPVNL
LLALFTVSLFTVGYSCANTEGCLPLNLSYCHKEQNLQETNSKACSCVETNNLL
FELLRLPGRIVLQALLITLSVSGSLPAYTPAAKKNKDFSLQPLIFTSLLILVYTS
FTQMFPLPGTISVAVYGGFSALKRRGRDSSSSSRKKLTDKTHSGAENVESSLT
DTGLTALANGFPRIENSLIMCVNSVGLCKSLQKTSLSKLSLDLQGVYDQGLAAV
GKFCQLEELNRFCEGLTDYGVLDLVVGCSSKLSIGVAAKAITDLSLEAVGSHCK
LLEVLVLDSEYIDHKGIAVAGQCHRLNKLQCVSVTVAFVAAVGLCTSLERLALY
SFQHDTKGMRAIGKSGKKLDLTSDCYFVSCKGLAIAHAGCKLERVINGCHNIG
TRGIEAIGKSCPLRELALYLCORIGNSALQIEIGKLYLKAGTDFHRTQNTGMDPLAEL
GECPPMLKDLVLSHCHHITDNLNHLVQCKLLETCHWVTCPGITSAGVATVVSSEPH

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.

REFERENCE AUTHORS TITLE

1 (bases 1 to 1589)
Graham, R.A. and Thornburg, R.W.

DNA Sequence of UDP Glucose:Indole-3-acetate
Beta-D-Glucosyltransferase from Arabidopsis thaliana (Accession No. U81293). (PGR97-044), 1004 (1997)

JOURNAL REFERENCE AUTHORS TITLE

2 (bases 1 to 1589)
Thornburg, R.W. and Graham, R.A.

Direct Submission
Submitted (06-DEC-1996) Biochemistry and Biophysics, Iowa State University, Ames, IA 50011, USA

FEATURES source

1..1589
Location/Qualifiers

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

26..1589

/gene="iaglu"

26..1450

/gene="iaglu"

/EC_number="2.4.1.121"

/function="transfers the glucose moiety from UDP-glucose

to indole-3-acetic acid to form ester linked IAA-Glc"

/note="IAA-glucose synthase"

/codon_start=1

/product="UDP-glucose:indole-3-acetate

beta-D-glucosyltransferase"

/protein_id="AAB58497.1"

/db_xref="GI:2149127"

/translation="NANNNSPTGPHELFVTPAOGHINPSLELAKRLAGTISGARV
TFAASISANRRWFSTENVPETLIPATSDGDDGFKSSAYSOKSDODATGNFWSMR
RRKGTLELIDNNRQNNPFCVVYITILLTVAELAREFLPALLWOPVTFVIF
YHFNFYEDAI SEMANTSSSIKLSPLLPVDRIPSFIVSSNYAFLLPFAFREQDS
LKEEINPKLIIINTFOLEPEAMSSVPDNFKIYPVPGPLTLRTDFSSRGEYIEWLDTKA
DSSLVVSFETLIVLSKKQLVELCKALIQSRPLWITDKSYRNKDEDEKEDCIS
SSKSPDEIDGMVVCDDQFVNLNHSIGCFVTHGWNSTLESLSVVPVVPQWMDQ
MTNAKLEDCWKTGYRVMKEKEEGVVVDSEIRRCIEEVMEDKAEPRGNATRKD
LAAEAREGGSFNLKAFVDEHM"

polyA_signal

1568..1575

/gene="iaglu"

polyA_site

1588..1589

/gene="iaglu"

/note="24 A nucleotides"

BASE COUNT 433 a 342 c 375 g 439 t

ORIGIN

Query Match 12.3%; Score 259.8; DB 8; Length 1589;

Best Local Similarity 55.9%; Pred. No. 4.6e-42;

Matches 672; Conservative 0; Mismatches 447; Indels 84; Gaps 6;

QY 556 tcattcaaaagaccctgtattttacatgtccaaacttcgttaactgtggaagtaactgt 615

DB 292 TCGTCAAGACCCACTGGAACACTCATGCTGAGATGACACGCTGGCAAGACACT 351

QY 615 caaaaaaataattcactgctctgtaaaatggacagcctataaacttgcctacttactc 675

DB 352 AACCGAACTAATCGAAGATAACCGGAAACAAACAGGCCCTTTTACTTCGTGGTTTAC 411

QY 676 catttctctctgggcagcagaggtagcaactgaagtcaactcctctctctctctct 735

DB 412 GATTCTCTCACTTGGGTGCTGAGCTAGCGCTGAGTTTCATCTTCCTCTGCTCTCT 471

QY 736 ttggagtcacacacacaaatattggacatatattacttcaacttcttcgatgataaaaa 795

DB 472 TTGGTCCAAACGACGATACGCTCTCCCAATTTTACCAATTTCAATGGCTAGGAAGA 531

QY 796 agctatggctaatgaatccaatgatccaaattggtccattcaacttccgggcttccact 855

|||||

Db 532 TGCAAATCTCAGAGATGGCTAATACCCCCCTCTAGTTCTATTAAATACCTTCTCTGCCACT 591

QY 856 actggaacatcgagatcttctcttcatttttacttctctatggtgcnaaaggagctcttcg 915

Db 592 GCTTACTGTGCGTGTATATTCCTTCTTCTTCTCTCTTCCAAATGCTACGCGTTCTT-- 649

QY 916 agtgcacttccaccattcaagaatgacacacattagatgctgaaacacactctctaa 975

Db 650 -----CTACCCCGCTTCGAGACAGATGATTACACTGAAGGAAGAAATAAACCCCTAA 702

QY 976 gatcttctgtaatacatcttgatgaattagagctgagcctcaatgcaatgaaggcta 1035

Db 703 GATCCTCATCAACTTTCACAGAGCTTGAGCAGAGCATGAGCTCGTTCGAGATAA 762

QY 1036 taagtcttgaatgagcagcttgattctctctctctctctctctctctctctctctct 1095

Db 763 -----TTTCAAGATTGCTCCTGCTCGTTTACTTAAGCTTGAGACGGA----- 807

QY 1096 agatgcttctattggtggtgatcttttccaaatccaaatgactataggaatggttaaa 1155

Db 808 -----TTTTTCGAGTCGCGTGAATACATAGAGTGTGTTGGA 843

QY 1156 ctcaaaagcnaattcatcagttgtttatatctcttttgggagctcaatgaatccactcat 1215

Db 844 TACTAAAGCGATTGCTGCTGCTTTATGTTTCTGCGGACGCTTCCGCTGTGTGAGCAA 903

QY 1216 tagccaaatggaggagagatacaaaaggggttgatagacataggaagggcgttttttatgggt 1275

Db 904 CAACACAGCTTGTGGAGCTTTGTAAGCGTTGTATACAAAGTCGGAGACCATCTTCTGTGGGT 963

QY 1276 gataaagaataaagcaaggaaggaaggaataaaaaagcttg----- 1322

Db 964 GATTACGGATAAGTCTGACAGAAATAAAGNAGATGAGCAAGAGGAAGAGATTGCAT 1023

QY 1323 -----gtgtattgaagaattggaaaaaataggaaaaaatagttccatggtgttccaaact 1377

Db 1024 AAGTAGTTACAGAGAAGAGCTTCGATGAGATGAGAAATGGTGGTTTCATGCTGTGATCAATT 1083

QY 1378 ggaagttctaaacacatctctttagagagttgtttctctctctctctctctctctctcgc 1437

Db 1084 TAGGCTTTTGAATCATAGATCGATAGTGTGTTTCGTGACGCTTCCGGGTGGAACCTCTAC 1143

QY 1438 cttagagagtttagctgtgagtgccagttgtggtggttctctctctctctctctctctct 1497

Db 1144 GCTGGAGAGCTTGGTTTTCAGGAGTTCGGTGGTGGCTTTCGCGCAATGGAATGATCAGAT 1203

QY 1498 gacaaatgccaaacaaagttaagatgtgtggaagagtgagtaagagtgga----- 1547

Db 1204 GACGAACGGAAGCTTTTAGAGATTCTTGGAAAAACAGCTGTAAGAGTGTAGGAGAGAA 1263

QY 1548 --gaataaatgaagatggtgtgtggaagtgaggaatacaaaagggtgattggaatggt 1605

Db 1264 GGAAGAAGAAGAGTTGCTGTGCTGATAGTAGAGAGATACGCGCGTGCATTTGAGGAAGT 1323

QY 1606 aatgagtgag 1665

Db 1324 TATGGA-----AGACAAAGCGGAGGAGTTAGAGCAATGCCACGAGGTGGAAGGATTT 1377

QY 1666 gactagagaagctgtgaggaagagtggtggtggtggtggtggtggtggtggtggtggt 1725

Db 1378 AGCGCGGAGGCTGTGAGAGAAGAGGCTCTCTCTTAAATCATCATCAAGGCTTTTGTGCGA 1437

QY 1726 tga 1728

Db 1438 TGA 1440

RESULT 12

AC005106/c

LOCUS

DEFINITION

AC005106

AC005106 84203 bp DNA

Genomic sequence for Arabidopsis thaliana BAC T25N20 from

Chromosome 1, complete sequence.

PLN

16-DEC-1999

T25N20 from


```

VERSION AC005106.2 GI:6587719
KEYWORDS
SOURCE HTG.
ORGANISM Arabidopsis thaliana
           thale cress.
REFERENCE Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H.,
AUTHORS Conway, A., Conway, A., Kutz, D., Ojima, S., Shen, Y., Toriumi, M.,
           Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and
           Ecker, J. R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T25N20 from
       Chromosome 1
JOURNAL 1 (bases 1 to 84203)
REFERENCE Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H.,
AUTHORS Conway, A., Conway, A., Kutz, D., Ojima, S., Shen, Y., Toriumi, M.,
           Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and
           Ecker, J. R.
TITLE Direct Submission
AUTHORS Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th Street and
           Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 84203)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center,
           Department of Biology, University of Pennsylvania, 38th Street and
           Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 84203)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center,
           Department of Biology, University of Pennsylvania, 38th Street and
           Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 5 (bases 1 to 84203)
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C.,
           Khan, S., Kim, C., Alt, F. H., Bei, B., Chin, C., Chlou, J., Choi, E.,
           Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howling, B., Koo, T.,
           Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S.,
           Mukharshy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J.,
           Southwick, A., Taveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R.,
           Federspiel, N., Theologis, A. and Ecker, J.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center,
           Department of Biology, University of Pennsylvania, 38th and
           Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT On Dec 16, 1999 this sequence version replaced gi:3366536.
FEATURES
       source
           1..84203
           /organism="Arabidopsis thaliana"
           /db_xref="taxon:3702"
           /chromosome="1"
           complement(join(1491..2159,2389..2452,2568..2703,
           3061..3149,3231..3277,3369..3457,4022..4139,4246..4353))
           /note="5' end of hypothetical protein YUP8H12.1; overlaps
           T7 end of YUP8H12"
           /codon_start=1
           /evidence=not_experimental
           /product="T25N20.1"
           /protein_id="AAC80580.1"
           /db_xref="GI:3935137"
           /translation="MGKKEQDRHVSVESDDKVEAVLHLRKHSLPTLTKQKFCNRAC
           VGRFIRKQDNVKAQKSLRSLSSRLSIADEFTAEGLAVVAGLDDECR
           PVLVFRKQDKLTKOLTRLVVTFLEVAISTMSRNVQFVILFDATAAAVLLVTL
           IIVLFSFNLVPLLILCHTAHLPEFFFLYKTNPKLLOISLSPHEHTCDQNSRRG
           IRTFVDTLSTWIMVQDFSDYDDFSSYSFVSRLAFDTLSLKTDKIGSCASS
           RFAFTVSDGLDVKPCLTUTDTSKLGHTNCAYISPLNARSFSFSPAPARSFPFG
           GPRFFASTPMPARTDRHSIGLPRIPRPSFPFQSPAIFFRSHVSKSEKPRDS
           FYQFLFYRRPYDEMTYRKRPLPGLGVISVTSQIRRRHVSLSORF"
           join(9304..11268,11387..11513,11607..11692,12103..12341,
           12698..12784,13257..13356,13451..14263)
           /note="unknown; similar to ESTs dbj|C26511 and
           gb|A1054650"

CDS
       1..2159,2389..2452,2568..2703,3061..3149,3231..3277,3369..3457,4022..4139,4246..4353
       /evidence=not_experimental
       /product="T25N20.5"
       /protein_id="AAC80584.1"
       /db_xref="GI:3935141"
       /translation="MMMDPPPTKEFALPVYVYVMDMKRCVPDDYDARRVGCPIKRIIL
       RKSVCGRITCGPGTAGQVDNFITHLSTREHKLKRRFTNPKGHLAN"
       complement(join(17577..18026,18231..18796,18981..19410))
       /note="similar to RNA polymerase II elongation factor
       homolog (S59773)"
       /codon_start=1
       /evidence=not_experimental
       /product="T25N20.4"
       /protein_id="AAC80583.1"
       /db_xref="GI:3935140"
       /translation="MGFLVAVMNSPLVHHMKSFKQCNELKRGQTSLSLDRRGF
       LKCVGASVFMTATIEFSGAQASEEKDEGEVYGAFKTLDFPNERTKSGKELPVEL
       LKSAREVVKTRSLKFNKPNKDAKRSDAASAEINDYLSNWRGQTVAGBESVVEL
       ENVIILAKFYSKAGSAPLPDEYKTEILDLANKAEFL"
       join(16409..16661,16979..17133,17234..17258,17346..17362)
       /note="unknown; similar to EST emb|Z26799"
       /codon_start=1
       /evidence=not_experimental
       /product="T25N20.3"
       /protein_id="AAC80582.1"
       /db_xref="GI:3935139"
       /translation="MGFLVAVMNSPLVHHMKSFKQCNELKRGQTSLSLDRRGF
       LKCVGASVFMTATIEFSGAQASEEKDEGEVYGAFKTLDFPNERTKSGKELPVEL
       LKSAREVVKTRSLKFNKPNKDAKRSDAASAEINDYLSNWRGQTVAGBESVVEL
       ENVIILAKFYSKAGSAPLPDEYKTEILDLANKAEFL"
       join(16409..16661,16979..17133,17234..17258,17346..17362)
       /note="unknown; similar to EST emb|Z26799"
       /codon_start=1
       /evidence=not_experimental
       /product="T25N20.2"
       /protein_id="AAC80581.1"
       /db_xref="GI:3935138"
       /translation="MEGGSGGERSIASNTPSLKEKSDGVCRTVSSSTETRONYQKR
       SRMVYDSSESDMKPPPSRGVDKTLGAKKEFKRDRVHEHDNGVVRNNSPAG
       SPWKANKLIDIEFDEYDGFSAANLKRFPONGSVGVRGSRSPASRVSDSVSGSGR
       EQLFRRRTFYNGTCSASSDSSSDSPKRVGGINGVLKYNKNTKTLAASI
       NPDRAETIERRPSSKKAORRNVVPPFRKSNVNNSSESDSVKRNKTKKRSKSEIS
       RPKKFNKTSKVPFELVNDVREERGGGDKORLRIKGLMDAGMTIDYKP
       RNNQSLDAVYVNSPGTAYNSIIKAYDALLKQKDEGVARPKRDTAAVSVSEIIVN
       KLARAKTRSEMTKKWQNSGSDSENKSEGAYTDTSEIRIRSSIKLGSKTKKGR
       NGADMDLHKRSKSLYNNARPCGSDSHYHGRKTKKIGRCTLIVRSKDKKNPAI
       NGFNYSGKRTLNLWLGESVOLRQVYRRRGAQVMEGLWIREGHCDCCKSKIL
       TVSPFIHAGSKSQPFONVLESGLAOCVQRANMOKDATNALHOLDVDDPN
       DDACGICGGDGLICDQCPSTYHONCLGMQVLPSCDWHCPNCTKCECAAASVSGKD
       GNFISLSCGMERYHOLCLNDFAHKVYQFGSASSFCGKCLEFKLGTGZVWTE
       IEGYSWLIHRVDTSDTNSQMAKRIENNSLAVGLAMDECFPLVDRRSQVLLI
       RNVLYNCSGNFNRYTGFYTAILERGETIISAASLRFHGMQLAEMPFITGTHYRRQ
       GMRCLFPAIESAMRSLKVKELVPAIPDFLHAWTNGFPTPLDSDVVRKEMSLNTLV
       FPGIDMLQKPLLIHEENIIAPAAAGDAMISEVETEKSEFTSSVEIGPYAVEGDEFVAD
       AANCYKDILASDEONILVSVETAMGTICPKDLSLRFHFRGEEGSISSSPCQITLKSQT
       KVLGHICDDTCSSCEDGLTDNVVEADASILSOEIOQASAFKVENLISLSIGRSS
       DLSSISQVKSQESNNLDGVPSCDYNILVPGAKLDSKDDAFACFL"
       complement(join(14820..14957,15038..15420,15643..15721))
       /note="hypothetical protein"
       /codon_start=1
       /evidence=not_experimental
       /product="T25N20.1"
       /protein_id="AAC80585.1"
       /db_xref="GI:3935142"
       /translation="MLNTYSILTLTKTKTPRVMMKNHPLCPSPSLTKPSSSPSGAT
       SDDPNRPSILLINNENLYDSSAAHRLSKPLIHDEVPSTSTATSTAAANSSSSA"

```


`/gene="F3F20.2"`
`/note="Hypothetical protein"`
`/codon_start=1`
`/protein_id="AAD30611.1"`
`/db_xref="GI:4836909"`
`/translation="MESNGGGEVRRNLYVFLSRSGHVDHPLLRVHLLSRNGVSEILR`
`DKVKLADARGDAMPADPASMCKRPKYNKVWQDLDDDDLTPTSDNEYVLKSGEILL`
`SSPKEDYPNVEKAWYTRNGIDAEKQLKLTSEKIQKESPFCSGORSTATTSVIT`
`EESTTNEEGVFLMKQGGPTSGQDSTNGSDNGVESGRPSVSTTSSSSYIKNKS`
`SVSRVASHLRNKKDGLDNTDVAIVLPNKSRSAGFAPWEDERRYYQHQHNARKSF`
`ENWASIMKNETIEFCPKVPKAPSCNAPLCSQCCKLPEKMKHSHMKLCRCGMKNSA`
`NDGATISNTNPKPQRCRINPNPLGHQRVLTTLKE"`
`complement(47118, .50318)`
`/gene="F3F20.3"`
`/complement(join(47118, .48467, 48761, .49149, 49186, .49302,`
`49523, .49979, 50064, .50318))`
`/gene="F3F20.3"`
`/note="similar to Na/H antiporter proteins; N-terminal`
`half of protein is similar to Na/H antiporter proteins"`
`/codon_start=1`
`/protein_id="AAD30610.1"`
`/db_xref="GI:4836908"`
`/translation="MSSGAGLVNTPNPDIEESRGCIVCYDOSLLPEKREKQGWESG`
`STLASSLPFIIOTFANLNVLYRLTYLTPPLPPVAQILCGLLFSPVLNTRFTI`
`AVHYEYRPTMLEVLFANLALYFLIFGLMGLRMVRLTELKPVILATPTGLVALPVG`
`AFELYLPNGHDKLIISGCPNVSVALACTNFPDLARILADLKLRSDNGRTAMCAALY`
`GDVLCVLLVGFSGFSKSGTGNKMPFVITTAIFVLCLIFVIRPGIAWILFAKTYKA`
`GGVYCGLLITDAGVSHSTGAFGLPSIHDDIIRNIEERLHDFSLGILMPLFYIIC`
`GLRADIYNGLOTRDKFNWVYVSCSEFLKIVTVITSLFMHIMPRDAFAGALMNTKG`
`TLSVLVLMQRDTKALDSPMYTHHTIALLVMSLVBPFLAAYKPKKLARHKRTVG`
`KIKGETELRVLCACVHLPVNSVITNLQVSNATKQSPISVFAIHLVELTGRTTASLI`
`MNDECKPRANFSDRVRAESDQIAETFEAMENVNDAMTQVITAVSPYATMHEDICVLA`
`EDKRCVFTILPVHKILTPDGRMGRENSSHAETINQVLSHAPCSGILVDRCGMVRSE`
`SPGRSEKREVAFLVFGDPDRREALSVARWYGOHVILKTVVRYVFGREALISSGKA`
`AEYEREKQVDDCIETEFNFKTMNDSSVKYIEKVVNDGQDTIATIREDDNNSVDLYVV`
`GRGYNDSPTVAGLMDSSSPLEGTIDGLTULASNFTNHAASVLVIOQYSATKRQAQVTA`
`AAATVMVGAACVGTGNNLESAGSGDAKMTDRDAEPFMKMSDEDEDEDEHQYGIHR`
`53195, .53210`
`/gene="F3F20.4"`
`join(51395, .52184, 52258, .53210)`
`/gene="F3F20.4"`
`/note="similar to hexosaminidase; Similar to`
`hexosaminidase, gi|1170249, gi|1076942 and others"`
`/codon_start=1`
`/protein_id="AAD30612.1"`
`/db_xref="GI:4836910"`
`/translation="MLTUSKPHVILIPFLTLLSPLFSIALPINIWPKPRFLSNWQ`
`HKAIALSNPTILAPHEQYLSASVYRHNLIIRSENTPSLISYPVKLMKRYTLNLYVT`
`VTDFLPLHHGVDEYSKLSIPIGSFSAAHLAHSAWAMRGLETFSPQWIGTSPDLCP`
`VGIYIQDPLRGHRLVDITSRNYGVDDIMRTIKAMSAKILNVPFHMTITDQSPFLP`
`LPERSLAAGSLGSDMYTTPEDYSKIVQYGEHGVYVLPDELDTGTHGSGNGEAYPELV`
`VTCAHFWPAPKASWEERLASPEGTQGNLPSPKTIETVKNYQIDVWQVPPFSFPHGG`
`GDEVTPGCMKTDPALNPSLGGGLTQLLEKYNSTLPYVYISQNTVYVWEDVLDAQ`
`IKADPSLPKPEHTLIQTWNNPENTKRIAAVGYRVIVSSSEFYIYDCDGHGFGFLNDIS`
`YDQKSGGSGSCAPKPTWQSTINTYDITADGILLNEERKELVIGGEVALWSEQADSTFSL`
`RLMRPRASALAESLWSGNDRGEYVRCRCGEAVDLRLNLRVYRWVRRKRGIGASPIQIPLWCLKN`
`PGMCNTVHGALQDQ"`
`complement(51395, .53210)`
`/gene="F3F20.5"`
`/complement(join(51395, .52184, 52258, .53210),`
`53685, .55199)`
`/gene="F3F20.5"`
`53685, .55199)`
`/gene="F3F20.5"`
`/note="Hypothetical protein"`
`/codon_start=1`
`/protein_id="AAD30626.1"`
`/db_xref="GI:4836924"`
`/translation="MSVVRVPRVLTPLLSQLLKQNPVTALKLFEAEKERFPSPYGH`

Query Match 11.0%; Score 232; DB 1; Length 476;
Best Local Similarity 99.6%; Pred. No. 1.8e-36;
Matches 243; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaatgtgagttagctcactcattagggcaccgccaggtttacac 60
|||||
Db 300 AGTGAGCGCAACGCAATTAAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACAC 241
|||||

QY 61 ttatgcttcggctcgatgtgtggaattgtgagcggaataacaatttcacacagga 120
|||||
Db 240 TTTATGCTTCGGCTCGATGTGTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 181
|||||

QY 121 aacagctatgaccatgattacgccaagctcgaaattaacccctcactaaagggaacaaaaa 180
|||||
Db 180 AACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAG 121
|||||

QY 181 ctgagctcca-cgcggtgcgccgctctagaactagtggatcccccgggctcagga 239
|||||
Db 120 CTGAGCTCCACCCGCGTGGCGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA 61
|||||

QY 240 ttcc 243
||||
Db 60 TTCC 57

Search completed: August 1, 2000, 13:51:48
Job time: 58570 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 18:19:33 ; Search time 152.01 Seconds
(without alignments)
1904.842 Million cell updates/sec

Title: US-09-147-955-11
Perfect score: 2105
Sequence: 1 agtgagcgcaacgcaattaa.....acaaaaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/PTCUS_COMB.seq.*
7: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231	11.0	7287	4	US-08-659-206A-1
C 2	230.6	11.0	3699	5	US-08-646-538-6
C 3	227.8	10.8	3792	4	US-08-992-334-1
C 4	227.8	10.8	3792	5	US-08-302-752-1
C 5	227.8	10.8	5234	4	US-08-992-334-2
C 6	227.8	10.8	5234	5	US-08-302-752-2
C 7	227.8	10.8	6722	4	US-08-992-334-3
C 8	227.8	10.8	6722	5	US-08-302-752-3
C 9	224	10.6	2885	2	US-08-471-496-1
C 10	224	10.6	2885	3	US-08-894-840-1
C 11	211.6	10.1	10281	4	US-08-816-155B-1
C 12	193.8	9.2	2415	1	US-07-885-970A-26
C 13	193.8	9.2	2415	1	US-08-298-687A-26
C 14	193.8	9.2	2415	1	US-08-298-829-26
C 15	192	9.1	14311	6	PCT-US96-06053-1
C 16	192	9.1	14311	6	PCT-US96-06053-7
C 17	188	8.9	4435	4	US-08-792-824-1
C 18	187.2	8.9	801	4	US-08-975-316-50
C 19	187.2	8.9	949	3	US-08-713-000-4
C 20	187.2	8.9	949	4	US-08-975-316-4
C 21	184.8	8.8	5534	2	US-08-452-267-3
C 22	184	8.7	5707	4	US-08-472-809B-8
C 23	184	8.7	6345	4	US-08-472-809B-7
C 24	183	8.7	4539	1	US-08-119-512-1
C 25	183	8.7	4539	2	US-08-488-015B-1
C 26	182.4	8.7	4525	3	US-08-613-861-2

27	182.2	8.7	7228	4	US-08-850-049-128	Sequence 128, App
28	182.2	8.7	7228	4	US-08-850-049-129	Sequence 129, App
29	182.2	8.7	7228	4	US-08-050-478-128	Sequence 128, App
30	182.2	8.7	7228	4	US-08-050-478-129	Sequence 129, App
C 31	181.8	8.6	4164	1	US-08-204-675-1	Sequence 1, Appli
C 32	181.8	8.6	4164	3	US-08-660-754-1	Sequence 1, Appli
C 33	181.8	8.6	4164	4	US-08-796-364-1	Sequence 1, Appli
C 34	181.8	8.6	4164	6	PCT-US95-02520-1	Sequence 1, Appli
C 35	181.8	8.6	4933	1	US-08-204-675-2	Sequence 2, Appli
C 36	181.8	8.6	4933	3	US-08-660-754-2	Sequence 2, Appli
C 37	181.8	8.6	4933	4	US-08-796-364-2	Sequence 2, Appli
C 38	181.8	8.6	4933	6	PCT-US95-02520-2	Sequence 2, Appli
C 39	181.2	8.6	535	2	US-08-463-115-57	Sequence 57, Appli
C 40	181.2	8.6	535	2	US-08-463-388-57	Sequence 57, Appli
C 41	180.8	8.6	5178	3	US-08-474-169-2	Sequence 2, Appli
C 42	190.8	8.6	6206	3	US-08-474-169-3	Sequence 3, Appli
C 43	179.2	8.5	3881	1	US-07-924-048A-6	Sequence 6, Appli
C 44	179.2	8.5	4792	3	US-08-596-300A-2	Sequence 2, Appli
C 45	179.2	8.5	5042	3	US-08-596-300A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-659-206A-1
; Sequence 1, Application US/08659206A
; Patent No. 5922685
; GENERAL INFORMATION:
; APPLICANT: Rakhmievich, Alexander
; TITLE OF INVENTION: IL-12 Gene Therapy of Tumors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,206A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 110229.91144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Plasmid DNA"
; IMMEDIATE SOURCE:
; CLONE: pWRG3169
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..628
; FEATURE:
; NAME/KEY: IDNA
; LOCATION: 629..810
; FEATURE:
; NAME/KEY: CDS

LOCATION: join(953..1258, 1332..1673)
OTHER INFORMATION: /product= "p35 gene product"
FEATURE:
NAME/KEY: polyA_site
LOCATION: 1797..2024
FEATURE:
NAME/KEY: promoter
LOCATION: 2110..2737
FEATURE:
NAME/KEY: IDNA
LOCATION: 2738..2919
FEATURE:
NAME/KEY: CDS
LOCATION: 2983..3990
OTHER INFORMATION: /product= "p40 gene product"
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4075..4306
US-08-659-206A-1

Query Match 11.0%; Score 231; DB 4; Length 7287;
Best Local Similarity 99.6%; Pred. No. 2.5e-43;
Matches 242; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 agtgagcgcaacgaataatgtgagttagctcactcattagggcaccacaggttttacac 60
DB 7024 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAGCTAGGACACCCAGGCTTTTACAC 7083
QY 61 ttatgcttcgggctcgatgtgtgtggaattgtgagcgataacaatttcacacagga 120
DB 7084 TTTATGCTTCCGGCTCGTATCTGTGTGGAATCTGAGCGGATAACAATTTTCACACAGGA 7143
QY 121 aacagctatgacctgattagcgaactcgaaattcaccctcactcaaaaggaacaaag 180
DB 7144 AACAGCTATGACCATGATAGCGCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAG 7203
QY 181 ctggagctccca-cgcgggtggcgccgctcagaaactagtgagtcaccccgggctgcaggaa 239
DB 7204 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGTGATCCCGCGGCTGCAGGAA 7263
QY 240 ttc 242
DB 7264 TTC 7266

RESULT 2
US-08-646-538-6/c
Sequence 6, Application US/08646538
Patent No. 6027881
GENERAL INFORMATION:
APPLICANT: Pavlakis, George N.
APPLICANT: Gaitanaris, George A.
APPLICANT: Stauber, Roland H.
APPLICANT: Vournakis, John N.
TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,538
FILING DATE: No. 6027881 yet assigned

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 015280-249000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3699 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..3699
OTHER INFORMATION: /note= "pbsgpp"
US-08-646-538-6
Query Match 11.0%; Score 230.6; DB 5; Length 3699;
Best Local Similarity 98.0%; Pred. No. 2.5e-43;
Matches 244; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 1 agtgagcgcaacgaataatgtgagttagctcactcattagggcaccacaggttttacac 60
DB 1678 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAGCTAGGACACCCAGGCTTTTACAC 1619
QY 61 ttatgcttcgggctcgatgtgtgtggaattgtgagcgataacaatttcacacagga 120
DB 1618 TTTATGCTTCCGGCTCGTATCTGTGTGGAATCTGAGCGGATAACAATTTTCACACAGGA 1559
QY 121 aacagctatgacctgattagcgaactcgaaattcaccctcactcaaaaggaacaaag 180
DB 1558 AACAGCTATGACCATGATAGCGCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAG 1499
QY 181 ctggagctccca-cgcgggtggcgccgctcagaaactagtgagtcaccccgggctgcaggaa 239
DB 1498 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGTGATCCCGCGGCTGCAGGAA 1439
QY 240 ttcggttc 248
DB 1438 TTCGATCGC 1430
RESULT 3
US-08-992-334-1/c
Sequence 1, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435


```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; IMMEDIATE SOURCE:
; CLONE: pg+host4
; US-08-992-334-1

Query Match 10.8%; Score 227.8; DB 4; Length 3792;
Best Local Similarity 98.8%; Pred. No. 1.1e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaataatgtgagtgtagctcactcattaggcaccacccgggtttacac 60
|||
Db 3758 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 3699

QY 61 ttatgtcttcggcgctgtgtgtggaattgtgagcggaataacaaattcacagga 120
|||
Db 3698 TTTATGCTTCGGCTCGTATGTGTGGAAATTGTGAGCGGATAACAATTTACACAGGA 3639

QY 121 aacagctatgaccatgattacgccaaagctcgaaattacccctcactaaagggaacaaag 180
|||
Db 3638 AACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCCTCACTAAAGGGAACAAAAG 3579

QY 181 ctgagactcca-cgcggtgcccgcctctagaactagtcaccccggtcgaggaa 239
|||
Db 3578 CTGAGCTCCACCGGTCGCGGCTCTAGACTAGTGTGATCCCGCGCTGCAGGAA 3519

QY 240 ttc 242
|||
Db 3518 TTC 3516

RESULT 4
US-08-302-752-1/c
; Sequence 1, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; { MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-752-1

Query Match 10.8%; Score 227.8; DB 5; Length 3792;
Best Local Similarity 98.8%; Pred. No. 1.1e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaataatgtgagtgtagctcactcattaggcaccacccgggtttacac 60
|||
Db 3758 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 3699

QY 61 ttatgtcttcggcgctgtgtgtggaattgtgagcggaataacaaattcacagga 120
|||
Db 3698 TTTATGCTTCGGCTCGTATGTGTGGAAATTGTGAGCGGATAACAATTTACACAGGA 3639

QY 121 aacagctatgaccatgattacgccaaagctcgaaattacccctcactaaagggaacaaag 180
|||
Db 3638 AACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCCTCACTAAAGGGAACAAAAG 3579

QY 181 ctgagactcca-cgcggtgcccgcctctagaactagtcaccccggtcgaggaa 239
|||
Db 3578 CTGAGCTCCACCGGTCGCGGCTCTAGACTAGTGTGATCCCGCGCTGCAGGAA 3519

QY 240 ttc 242
|||
Db 3518 TTC 3516

RESULT 5
US-08-992-334-2/c
; Sequence 2, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maquin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match      10.8%; Score 227.8; DB 4; Length 5234;
Best Local Similarity 98.8%; Pred. No. 1.2e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtggcgcaacgaataatgtgagtagctcaactcattaggcaccacaggctttacac 60
|||||
Db 5200 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAGTATAGGCACCCGAGCTTTACAC 5141
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaatacaatttcacacagga 120
|||||
Db 5140 TTTATGCTTCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGA 5081
QY 121 aacagctatgacctgattacgcgaagctcgaaattaaacctcaactaaagggaacaaaag 180
|||||
Db 5080 ACACGCTATGACCATGATTACGCCAAGCGCGCAATTAAACCTCAGTAAAGGGAACAAAAG 5021
QY 181 ctggagctcca-cgcgggtggcgccgctctagaaactagtggaatcccccggtgcaggaa 239
|||||
Db 5020 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA 4961
QY 240 ttc 242
|||
Db 4960 TTC 4958

RESULT 7
US-08-992-334-3/c
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match      10.8%; Score 227.8; DB 4; Length 5234;
Best Local Similarity 98.8%; Pred. No. 1.2e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtggcgcaacgaataatgtgagtagctcaactcattaggcaccacaggctttacac 60
|||||
Db 5200 AGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAGTATAGGCACCCGAGCTTTACAC 5141
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaatacaatttcacacagga 120
|||||
Db 5140 TTTATGCTTCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGA 5081
QY 121 aacagctatgacctgattacgcgaagctcgaaattaaacctcaactaaagggaacaaaag 180
|||||
Db 5080 ACACGCTATGACCATGATTACGCCAAGCGCGCAATTAAACCTCAGTAAAGGGAACAAAAG 5021
QY 181 ctggagctcca-cgcgggtggcgccgctctagaaactagtggaatcccccggtgcaggaa 239
|||||
Db 5020 CTGGAGCTCCACCGCGGTGGCGCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA 4961
QY 240 ttc 242
|||
Db 4960 TTC 4958

RESULT 6
US-08-302-752-2/c
; Sequence 2, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match          10.8%; Score 227.8; DB 4; Length 6722;
Best Local Similarity 98.8%; Pred. No. 1.3e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagttgagctcactcattaggcaccgccaggtttacac 60
Db 6688 AGTGAGCGCAACGAATTAATGTGAGTTGAGCTCATTAGGCACCCCGAGGCTTTACAC 6629

QY 61 ttatgcttcggctcgatgtgtggaattgagcgataacaaatttcacagga 120
Db 6628 TTTATGCTTCCGGCTCGTATGTTGTGTAATGTGAGCGGATAAACAATTTACACAGGA 6569

QY 121 aacagctatgaccatgattacgcgaagctcgaataataaccctcactaaagggaacaaaag 180
Db 6568 AACAGCATATGACCATGATTACGCCAAGCGCGCAATTAACCCCTACTAAGGGACAAAG 6509

QY 181 ctgagactcca-cgcggtggcgccgctctagaaactagtgatcccccggtgcaggaa 239
Db 6508 CTGAGACTCCACCGCGGTGGCGGCTCTAGAACTAGTGGATCCCGCGGCTGCAGAA 6449

QY 240 ttc 242
Db 6448 TTC 6446

RESULT 8
US-08-302-752-3/c
; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match          10.8%; Score 227.8; DB 5; Length 6722;
Best Local Similarity 98.8%; Pred. No. 1.3e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagttgagctcactcattaggcaccgccaggtttacac 60
Db 6688 AGTGAGCGCAACGAATTAATGTGAGTTGAGCTCATTAGGCACCCCGAGGCTTTACAC 6629

; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match          10.8%; Score 227.8; DB 4; Length 6722;
Best Local Similarity 98.8%; Pred. No. 1.3e-42;
Matches 240; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagttgagctcactcattaggcaccgccaggtttacac 60
Db 6688 AGTGAGCGCAACGAATTAATGTGAGTTGAGCTCATTAGGCACCCCGAGGCTTTACAC 6629

QY 61 ttatgcttcggctcgatgtgtggaattgagcgataacaaatttcacagga 120
Db 6628 TTTATGCTTCCGGCTCGTATGTTGTGTAATGTGAGCGGATAAACAATTTACACAGGA 6569

QY 121 aacagctatgaccatgattacgcgaagctcgaataataaccctcactaaagggaacaaaag 180
Db 6568 AACAGCATATGACCATGATTACGCCAAGCGCGCAATTAACCCCTACTAAGGGACAAAG 6509

QY 181 ctgagactcca-cgcggtggcgccgctctagaaactagtgatcccccggtgcaggaa 239
Db 6508 CTGAGACTCCACCGCGGTGGCGGCTCTAGAACTAGTGGATCCCGCGGCTGCAGAA 6449

QY 240 ttc 242
Db 6448 TTC 6446

RESULT 9
US-08-471-496-1
; Sequence 1, Application US/08471496
; Patent No. 5798223
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; APPLICANT: CAO, LIANG
; TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,496
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/02645
; FILING DATE: 01-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488-0830001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 719..2128
US-08-471-496-1

Query Match          10.6%; Score 224; DB 2; Length 2885;
Best Local Similarity 98.7%; Pred. No. 7.2e-42;
Matches 235; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgaattaatgtgagttgagctcactcattaggcaccgccggtttacac 60
Db 6688 AGTGAGCGCAACGAATTAATGTGAGTTGAGCTCATTAGGCACCCCGAGGCTTTACAC 6629
```

Db 207 AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTACTCTACTCTACTCTAGTGGCACCCTGAGGCTTTACAC 266
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaataacaatttcacacagga 120
Db 267 TTTATGCTTCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 326
QY 121 aacagctatgacctgattacgcgaactcgaattaaacctcaactaaagggaacaaaag 180
Db 327 ACACGCTATGACCATGATTACGCCCAAGCTCGAATTAACCTCTACTTAAGGGAACAAAAG 386
QY 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccccgggctgcagg 237
Db 387 CTGGAGCTCCACCGCGTGGCGNCCGCTCTAGAACCTAGTGTGATCCCGCGNCTGCAGG 444

RESULT 10
US-08-894-840-1
; Sequence 1, Application US/08894840
; Patent No. 5859200
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; APPLICANT: CAO, LIANG
; APPLICANT: ROSEN, CRAIG A.
; TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,840
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STERNE, ERIC K
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0830000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2500
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 719..2128
; US-08-894-840-1

Query Match 10.6%; Score 224; DB 3; Length 2885;
Best Local Similarity 98.7%; Pred. No. 7.2e-42;
Matches 235; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaaatgtgagttagctcactcactcattaggcaccgccagctttacac 60
Db 207 AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTACTCTACTCTAGTGGCACCCTGAGGCTTTACAC 266
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaataacaatttcacacagga 120
Db 267 TTTATGCTTCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 326
QY 121 aacagctatgacctgattacgcgaactcgaattaaacctcaactaaagggaacaaaag 180

Db 327 AACAGCTATGACCATGATTACGCCCAAGCTCGAATTAACCTCTACTTAAGGGAACAAAAG 386
QY 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccccgggctgcagg 237
Db 387 CTGGAGCTCCACCGCGTGGCGNCCGCTCTAGAACCTAGTGTGATCCCGCGNCTGCAGG 444

RESULT 11
US-08-816-155B-1/c
; Sequence 1, Application US/08816155B
; Patent No. 5990091
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.
; APPLICANT: MARTINEZ, HECTOR
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: PINCUS, STEVEN E.
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,155B
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KOMALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454310-2990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10281 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-816-155B-1

Query Match 10.1%; Score 211.6; DB 4; Length 10281;
Best Local Similarity 97.8%; Pred. No. 6.3e-39;
Matches 225; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 agtgagcgcaacgcaattaaatgtgagttagctcactcactcattaggcaccgccagctttacac 60
Db 4083 AGTGAGCGCAACGCAATTAATGTGAGTTAGTCTACTCTACTTATAGGCACCCAGGCTTTACAC 4024
QY 61 ttatgtctccggctcgatgtgtgtggaattgtgagcggaataacaatttcacacagga 120
Db 4023 TTTATGCTTCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTTCACACAGGA 3964
QY 121 aacagctatgacctgattacgcgaactcgaattaaacctcaactaaagggaacaaaag 180
Db 3963 AACAGCTATGACCATGATTACGCCCAAGCTCGAATTAACCTCTACTTAAGGGAACAAAAG 3904
QY 181 ctggagctcca-cgcggtggcgccgctctagaaactagtagtgcaccccccgg 229

Db 3903 CTGAGCTCCACCGGCTGGCGCGCTCTAGAACTAGTGTGATCCTGCCG 3854

RESULT 12

US-07-885-970A-26
; Sequence 26, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885.970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617.239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253.243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27.386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2415 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium barbadense
; STRAIN: Sea Island
; IMMEDIATE SOURCE:
; LIBRARY: EMBL SI
; CLONE: SIB12
US-07-885-970A-26

Query Match 9.2%; Score 193.8; DB 1; Length 2415;
Best Local Similarity 99.0%; Pred. No. 4.3e-35;
Matches 195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 agtgagcgcaacgcaattaatgtgagtgatgctcactcattaggtaccccccaggctttacac 60
Db 241 AGTGAGCGCAACGCAATTAAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 300
Qy 61 ttatgcttccggctcgatgtgtggaattgtgagcggtatacaaatccacagga 120
Db 301 TTTATGCTTCCGGCTCGATGTGTGTGGAATTGTGAGCGGATAAACAATTTACACAGGA 360
Qy 121 aacagctatgaccatgattacgcgaagctcgaaattacccctcactaaaagggaacaaag 180
Db 361 AACAGCTATGACCATGATTACGCCAAGCTCGCAATTAACCCCTCATTAAACGGGAACAAAG 420

Qy 181 ctggagctccacgcggt 197
Db 421 CTGAGCTCCACATGGT 437

RESULT 13

US-08-298-687A-26
; Sequence 26, Application US/08298687A
; Patent No. 5521076
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298.687A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617.239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253.243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27.386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2415 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium barbadense
; STRAIN: Sea Island
; IMMEDIATE SOURCE:
; LIBRARY: EMBL-SI
; CLONE: SIB12
US-08-298-687A-26

Query Match 9.2%; Score 193.8; DB 1; Length 2415;
Best Local Similarity 99.0%; Pred. NO. 4.3e-35;
Matches 195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 agtgagcgcaacgcaattaatgtgagtgatgctcactcattaggtaccccccaggctttacac 60
Db 241 AGTGAGCGCAACGCAATTAAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACAC 300
Qy 61 ttatgcttccggctcgatgtgtggaattgtgagcggtatacaaatccacagga 120
Db 301 TTTATGCTTCCGGCTCGATGTGTGTGGAATTGTGAGCGGATAAACAATTTACACAGGA 360
Qy 121 aacagctatgaccatgattacgcgaagctcgaaattacccctcactaaaagggaacaaag 180

Db 361 AACAGCTATGACCATGATTACGCAAGCTCGAAATTAACCTTCACCTAAAGGGAACAAAG 420
Qy 181 ctggagctccacgcggt 197
Db 421 CTGGAGCTCCACATGGT 437
RESULT 14
US-08-298-829-26
; Sequence 26, Application US/08298829
; Patent NO. 5620882
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; City: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,829
; FILING DATE: 19-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/885,970
; FILING DATE: 18-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2415 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium barbadense
; STRAIN: Sea Island
; IMMEDIATE SOURCE:
; LIBRARY: EMBL-SI
; CLONE: SIB12
US-08-298-829-26

Query Match 9.2%; Score 193.8; DB 1; Length 2415;
Best Local Similarity 99.0%; Pred. No. 4.3e-35;
Matches 195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 agtgagcgaacgcaataatgtgagttagctcactcattagggcaccacccaggctttacac 60
Db 241 AGTGAGCGCAACGCAATATATGTGAGTTAGCTACATCATTAGGCACCCAGGCTTTACAC 300

Qy 61 ttatgtcttcgctcgtatgtgtgtggaattgtgagcggataacaatttcacacagga 120
Db 301 TTTATGCTTCGGCTCGTATGTTGTGTGGAATTTGTGAGCGGATAACAATTTTCACACAGGA 360
Qy 121 aacagctatgacatgattacgcgaagctcgaaattaaacctcactaaagggaaacaaaag 180
Db 361 AACAGCTATGACCATGATTACGCCAAGCTCGAANTTAACCTTCACCTAAAGGGAACAAAG 420
Qy 181 ctggagctccacgcggt 197
Db 421 CTGGAGCTCCACATGGT 437
RESULT 15
PCT-US96-06053-1/c
; Sequence 1, Application PC/TUS9606053
; GENERAL INFORMATION:
; APPLICANT: Yale University
; TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR
; TITLE OF INVENTION: USES
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; City: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06053
; FILING DATE: 01-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6523-009-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14311 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 760..2028
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2092..2889
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2946..3635
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3774..5309
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5429..11758
PCT-US96-06053-1

Query Match 9.1%; Score 192; DB 6; Length 14311;
Best Local Similarity 100.0%; Pred. No. 1.8e-34;

	Matches	192;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	agtgagcgcaacgcaataatgtgagttagctcactcattagcgacccagcgtttacac	60							
Db	12296	AGTGAGCGCAACGCAAA'TTAATGTGAGTTAGCTCAGTTCATTAGGACCCCGAGCTTTACAC	12237							
QY	61	ttaatgtctcggctcgtatgtttgtggaattgtgagcgggatacaaatttcacacagga	120							
Db	12236	TTTATGTCTTCGCGGCTCGTATGTTGTGTGSAAT'TGTGAGCGGATTAACAATTTTCACACAGGA	12177							
QY	121	aacagctatgacatgatcagcccaagctcgaaatataccctcactaaagggaacaaaag	180							
Db	12176	ACACGATATGACCATGATTAGCCGCCAGCTCGAAATTAACCTTCCTTAAGGGAACAAAG	12117							
QY	181	ctggagactccac	192							
Db	12116	CTGGAGCTCCAC	12105							

Search completed: August 1, 2000, 18:20:49
Job time: 74498 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		ID	Description
		Match	Length		
1	750.5	30.3	471	1	IAAG_NA1ZE
2	553.5	22.4	453	1	UF0G_GENPR
3	435	17.6	459	1	ZOG_PRLU
4	432.5	17.5	487	1	UF0J_WANES
5	432	17.5	449	1	UF0J_WANES
6	431	17.4	346	1	UF02_WANES
7	417	16.8	471	1	UF01_NA1ZE
8	409	16.5	471	1	UF03_NA1ZE
9	408.5	16.5	433	1	UF0G_SOLME
10	404.5	16.3	454	1	ZOX_PRLU
11	404	16.3	471	1	UF02_NA1ZE
12	393.5	15.9	394	1	UF06_WANES
13	392.5	15.9	287	1	UF07_WANES
14	388	15.7	455	1	UF0G_HORVU
15	375	15.2	473	1	UF0G_PETHY
16	372.5	15.1	241	1	UF04_WANES
17	279	11.3	154	1	UF0G_VITVI
18	189	7.6	533	1	UD12_RAT
19	187.5	7.6	531	1	UD16_MOUSE
20	185	7.5	535	1	UD11_RAT
21	184	7.4	531	1	UD13_RAT
22	183	7.4	532	1	UD14_RABIT
23	183	7.4	534	1	UD14_HUMAN
24	182	7.4	535	1	UD11_MOUSE
25	181.5	7.3	533	1	UD12_MOUSE
26	180	7.3	533	1	UD11_HUMAN
27	179.5	7.3	430	1	OLED_STRAT
28	176	7.1	703	1	YK76_CAEEL
29	174.5	7.1	534	1	UD15_HUMAN
30	173	7.0	529	1	UD16_RAT
31	172.5	7.0	531	1	UG7E_CAEEL
32	171.5	6.9	534	1	UD13_HUMAN
33	171	6.9	525	1	UG7G_CAEEL

	Query Match	30.3%	Score 750.5;	DB 1;	Length 471;
Best Local Similarity	35.5%;	Pred. No. 2.7e-49;			
Matches 172;	Conservative	85;	Mismatches 188;	Indels 39;	Gaps 13;
QY	4	PHVILTPAAGHINPALQFAKNLVKMGIEVTFSTSYAQSRMD---	EKSILNAPKGLNF	60	
		: : : : : : : : : :			
DB	3	PHLVVVPFGGHNMPVQFAKSLAKSGVATLTVLTFTORTADYDAHPAMVEA-----	56		
QY	61	IPFSDGFDE-GFDSKDPVFMVMSQLRCGSETVKKIILTCSENGOPICTLIYSIFLPWAA	119		
		: : : : : : : : : :			
DB	57	--ISDGHDEGGFASGAAGVAYLEKQAAAAASLASIAVEARASDAFTCVVYDSVDKVL	114		
		: : : : : : : : : :			

```

QY 120 EVAREVHIPALLWSQPATILDIYFNFHGY-----EKAMANESNDPNWSIQLP-----GLP 171
DB 115 PVARMGLPVPFSTQSCAVSYVYHFSOGRVAVPPGAANDSGDGGAGAAALSAEAFGLP 174
QY 172 LLETRDLPFLPYGAKGSRVLPFPKELIDTDAETTPKILVNTFDELEPEALNAIEG 231
DB 175 EMERSELPSPFVDFHGPYPT--IAMQIKQF--AHAGKDDWLFNSFEETEVLGLTK 229
QY 232 Y-KRYGTGCLIP---SAFLGGNDPLDASFGDLPQNSNDYNEWLNSKPNSSVYISFGS 286
DB 230 YLKARAGPCVPLPTAGRTAGNRI--TYGANLVKPEDACTKMDTKPDRSVAVYVSFGS 287
QY 287 LNPSPISQMEISIKGLIDIGRPFLWIKENEGKEENKKGICIEELEKIG--KIVPWCS 344
DB 288 LASLGNQAQKEELARGLLAACKPELWVY---RASDEHQVPRYLLAEATATGAAMVVPWP 343
QY 345 QLEVLKHPSCGCFVSHCGWNSALESACGVPVVPFOWTQMTNAKQVEDYVWKSQVRVRI 404
DB 344 QLDVLHAPVAGCFVTHCGWNSLEALSFGVPMVAMALWTDQPTNARNVELAWGAGVRARR 403
QY 405 NED-GVVESEIKRICIELVMDGEGKEELRNKAKWELAREAVKEGSSHNKAKFIDD 463
DB 404 DAGAGVFLRGEVERCVRAVMDGGEASAAKKAAGEWRDRARAAPGCGSSDRNLDEFVQF 463
QY 464 YAKG 467
DB 464 VRAG 467

RESULT 2
ID UFGO_GENTR STANDARD; PRT; 453 AA.
AC Q96493;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
  3-O-GLUCOSYLTRANSFERASE).
OS Gentiana triflora.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Gentianales;
OC Gentianaceae; Gentiana.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PETAL;
RX MEDLINE; 96416435.
RA Tanaka Y., Yonekura K., Fukuchi-Mizutani M., Fukui Y.,
RA Fujiwara H., Ashikari T., Kusumi T.;
RT Molecular and biochemical characterization of three anthocyanin
RT synthetic enzymes from Gentiana triflora.;
RL Plant Cell Physiol. 37:711-716(1996).
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC GLYCOSYLATION REACTION: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; D85186; BAA12737.1; -
DR PFAM; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 453 AA; 5009 MR; BF738B0A2DA76C05 CRC64;

```

```

Query Match 22.4%; Score 553.5; DB 1; Length 453;
Best Local Similarity 29.7%; Pred. No. 1.9e-34;
Matches 143; Conservative 92; Mismatches 181; Indels 65; Gaps 14;

QY 5 HVLTTPFAQGHINPALQFAKNLVKMGIEVTF-----TSIVAQSRMDEKSLNAPK 56
DB 6 HVAVLAFPGFTHAARPLTLVNRLLAASAPDIIFSFSTSSITITFSPNL-----ISI 58
QY 57 GUNFIPFS--DGFDEGEDHNDP-----VFYMSOLRCGSETVKKILILTCSENGQIPITCL 110
DB 59 GSNIKPYAVMDGSPGEFVFGNPREPIEYFLNAPDNFDMKAMKAV---EDTGVNISCL 115
QY 111 YSIFUPWAAEVARVHPHSALLMSOPATILDIYF-----NPHGVKA-MANESNDPNW 163
DB 116 TDAFLWFAADSEKIGVWPVPTWTAASCSLCILHVYTDERSFAEDTAERAKETID--- 172
QY 164 SIQLPGLPLTRDLPFLPYGAKGSLRVALPFPKELIDTDAETTPKILVNTFDELEP 223
DB 173 --FIPGLSAISFSDLPBELIMEDSQSIFALTLMN-----GLKLKATAVAVNSPEEIDP 225
QY 224 EALNAIEG--YKFGYGLPILPSAFLGGNDPLDASFGDLPQNSNDYNEWLNSKPNSSV 280
DB 226 IITNHLRSTNQLNINIGPL-----QTLSSSIPPEDNECLKWLQTKRESSV 272
QY 281 YISFGSLMNPISQMEISIKGLIDIGRPFLWIKENEGKEENKKGICIEELEKIGIV 340
DB 273 YLSFGTVINPPNEMAALASTLESRKIPFLWSLRDEARKHLPEN----FIDRTSTFGKIV 328
QY 341 PWCSOLEVLKHPISLGCFSVSHCGWNSALESACGVPVVPFOWTQMTNAKQVEDYVWKSQV 400
DB 329 SWAPQLHVLENPALGVFTHCGWNSLTLESFCRVPIVGRPFDDQKYNARWVEDWKICV 388
QY 401 RVRINEDGVSESEIKRICIELVMDGEGKEELRNKAKWELAREAVKEGSSHNKAKF 460
DB 389 GVK--GGVFTTEITRYLELVL-FSDKGKEMRQNVGRLEKAKADAVKANGSSTRNFESL 444
QY 461 I 461
DB 445 L 445

RESULT 3
ID ZOG_PHALU STANDARD; PRT; 459 AA.
AC Q92SK5;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE ZEATIN O-GLUCOSYLTRANSFERASE (EC 2.4.1.203) (ZEATIN O-BETA-D-
  GLUCOSYLTRANSFERASE).
GN ZOG1.
OS Phaseolus lunatus (Lima bean) (Phaseolus limensis).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; euroids I; Fabales; Fabaceae;
OC Papilionoideae; Phaseolus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. KINGSTON; TISSUE=IMMATURE SEED;
RX MEDLINE; 99093528.
RA Martin R.C., Mok M.C., Mok D.W.S.;
RT *Isolation of a cytokinin gene, ZOG1, encoding zeatin
RT O-glucosyltransferase from Phaseolus lunatus.;
RL Proc. Natl. Acad. Sci. U.S.A. 96:284-289(1999).
CC -!- FUNCTION: MAY REGULATE ACTIVE VS. STORAGE FORMS OF CYTOKININS AND
CC COULD HAVE AN IMPACT ON SEED GROWTH. CAN ALSO USE UDP-XYLOSE TO
CC CATALYZES THE FORMATION OF O-XYLOSEZEATIN BUT AT MUCH LOWER
CC AFFINITY.
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + ZEATIN = UDP + O-BETA-D-
CC GLYCOSYLZEATIN.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----

```



```

RESULT 5
UFO1_MANES STANDARD; PRT; 449 AA.
AC Q40284;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 1 (EC 2.4.1.91) (UDP-GLUCOSE
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 1).
GN GT1 OR UGT73A1.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
OC Euphorbiaceae; Manihot.
RN [1]
RP SEQUENCE FROM N. A.
RC TISSUE-COTYLEDON;
RX MEDLINE; 95201291.
RA Hughes J., Hughes M.A.;
RT *Multiple secondary plant product UDP-glucose glucosyltransferase
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.*;
RL DNA Seq. 5:41-49(1994).
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS AND ROOTS, BUT NOT IN
CC LEAVES.
CC -1- DEVELOPMENTAL STAGE: MAXIMUM EXPRESSION IN COTYLEDONS THAT JUST
CC EMERGED FROM THE SEED COAT. LOW LEVELS IN HYPOCOTYLS AND
CC INCREASING LEVELS IN ROOTS THROUGHOUT THIS PERIOD OF DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X77459; CAA54609.1; -.
CC PFAM; PF00201; UDPGT; 1.
CC PROSITE; PS00375; UDPGT; 1.
CC TRANSFERASE; Glycosyltransferase; Multigene family.
CC SEQUENCE 449 AA; 50282 MW; A6738182F316DE75 CRC64;

Query Match 17.5%; Score 432; DB 1; Length 449;
Best Local Similarity 28.5%; Pred. No. 2.7e-25;
Matches 139; Conservative 84; Mismatches 189; Indels 76; Gaps 17;

QY 15 GHINPQAKNLVKMGIEVTFSTSIYAQSRMDEK-----SILNPKGLNFI----- 61
DB 2 GHLYSAVETAKLLLSRCHSLSIVLVFNNSVTSKYHNVDSQIASSNRLRFTYLPKDE 61
QY 62 ----PFSDGDFEGDHSKDPVYVYNSQLRKGSTVKKIILTCSENGQPITCLLYSIFLPW 117
DB 62 TGISFSFSLIKQRPYKESVMKITEF--GS-----SVESPLRVGFIVDMFCTA 108
QY 118 AAAYAREVHPSALLWSQPATILD--IYVFNHGYEKAMNESNDPNNSIOLPGLPLET 175
DB 109 MIDVANEFGVPSYIFYTSGAFLNFMHLVQKIHDEENPFTFNASDGLQVPLG----V 164
QY 176 RDLPSFLPYGAKGSLRVALLPPPKELIDTLDATETPKILVNTFDELEPEALNAGEYKPY 235
DB 165 NSFPKAMPPTAILS--KQWFPPLLE--NTRRYGEAKGVYINTFFFEASHAESFKDPPIY 220
QY 236 GIGPLIPSAFLGGNDPLDASFGGDLFQNSN-DYMEWLNKPNSSVVYISFGSLMNPISQ 294
DB 111

```

```

Db 221 PVGPI-----LDVRSNG-----RNTNQEIOMQLDDQPPSSVVFCLPSSNGSFSKQ 266
QY 295 MBEISKGLIDIGRPFLLWIK-----ENEKGEENKLL--GCIEELEKIGKIVPWS 344
DB 267 VKETACALEDSGHRFLWSLADHRAPGFLSPSYEDLQVLPPEGFLERTSGIEKVGNAP 326
QY 345 QLEVLKPSLGCFSYHCGWNSALESACGVPVAVFPQWTDQMTNAKQVEDYWKSGVRRI 404
DB 327 QVAVLAHPATGGLVSHSGWNSILESIFWGPVATWPNYAEQOQFNAPQM--VTELGLAVEI 384
QY 405 -----NEDG--VVESEEKRCIELVMGGEERKRNKAKKWLAREAVKEGSSHKNLK 458
DB 385 KMDYRNDSEGVTKQDIERGIRCLM-----KHSDRRKKVKEMSEKSRGALMEGSSYCWLD 441
QY 459 AFIDDVAK 456
DB 442 NLIKDMIK 449

RESULT 6
UFO2_MANES STANDARD; PRT; 346 AA.
AC Q40285;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 2 (EC 2.4.1.91) (UDP-GLUCOSE
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 2) (FRAGMENT).
GN GT2 OR UGT73A2.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
OC Euphorbiaceae; Manihot.
RN [1]
RP SEQUENCE FROM N. A.
RC TISSUE-COTYLEDON;
RX MEDLINE; 95201291.
RA Hughes J., Hughes M.A.;
RT *Multiple secondary plant product UDP-glucose glucosyltransferase
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.*;
RL DNA Seq. 5:41-49(1994).
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS, ROOTS AND LEAVES.
CC -1- DEVELOPMENTAL STAGE: MAXIMUM EXPRESSION IN COTYLEDONS THAT JUST
CC EMERGED FROM THE SEED COAT. LOW LEVELS IN HYPOCOTYLS AND
CC INCREASING LEVELS IN ROOTS THROUGHOUT THIS PERIOD OF DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X77461; CAA54611.1; -.
CC PFAM; PF00201; UDPGT; 1.
CC PROSITE; PS00375; UDPGT; 1.
CC TRANSFERASE; Glycosyltransferase; Multigene family.
CC NON_TER 1
CC SEQUENCE 346 AA; 38834 MW; 89D51FDF8E0D5F2 CRC64;

Query Match 17.4%; Score 431; DB 1; Length 346;
Best Local Similarity 32.8%; Pred. No. 2.3e-25;
Matches 121; Conservative 64; Mismatches 134; Indels 50; Gaps 13;

```

```

QY 120 EVAREVHPSALLWSOPATILD--IYYNFHGYEKAMANESNDPNWSIQLPGL--PLLET 175
D 7 DLAEFGIPSYIFASGGFLGFMLYVKIHDENFNIEFKSDTELIYPSLVNP-FPT 65
QY 176 RDLPSFLLPYGAGSRVALPFPKELIDTDAETTPKTLVNTFDELEPEALNAIEGKVEY 235
D 66 RILPSSILANKERGQLLATAKFRQ-----AKGIIVNTFLELESRALESFKVPPLY 116
QY 236 GIGPLPSAFLOGNDLDSFGDGLFQNSN-DYHNLNSKNSSVYVIFSGSLMNPISQ 294
D 117 HVGPI-----LDVKSOG---RNTHPEIMQWLDQPEGSVWFLFCGMSGSEDO 162
QY 295 MEETSKGLIDIGRPFLLWAKE-----NEKKEENKLL--GCIELEKIGKIYVWCS 344
D 163 LKEITVALENSGHRFLWSIRPPPPDKIASPTDYEDPRDVLPEGFLERTVAVCKVIGWAP 222
QY 345 QLEVLKPSLCGVSHCGNWSALESACGVVPVAFPPQWTDQMTNAKQVDEWVKSQVRYI 404
D 223 QVAVLAHPAIGGVSHCGNWSLESWFGVPIATWPMYAEQFNA--FEMVVELGLGVEI 280
QY 405 N-----EDG-VVSEBIRKRCIELVMDGGEKEELKNAKWKELAREAVKEGSSSHKLNK 458
D 281 DMGVRKESGIIVNSDKIERAIRKLM---ENSDEKRRKKVKEMREKSKNALIDGSSSFISLG 337
QY 459 AFIDDVAKG 467
D 338 DFIDKAMEG 346

RESULT 7
UFOL_MAIZE STANDARD; PRT; 471 AA.
ID UFOL_MAIZE AC P16167;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MCC ALLELE).
OS BZ1 OR UGT71A1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RA Furtek D., Schlefelbein J.W., Johnston F., Nelson O.E. Jr.;
RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL Plant Mol. Biol. 11:473-481(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88284304.
RA Ralston E.J., English J.J., Dooner H.K.;
RT "Sequence of three bronze alleles of maize and correlation with the
RT genetic fine structure.";
RL Genetics 119:185-197(1988).
CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13500; CAA31855.1; -.

```

```

DR EMBL; X07940; CAA30761.1; -.
DR PIR; S01052; S01052.
DR PIR; S08324; S08324.
DR MAIZEDB; 13885; -.
DR PFAM; PF00201; UDPGT; 2.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 471 AA; 48769 MW; 6234FD59219AF534 CRC64;

Query Match 16.8%; Score 417; DB 1; Length 471;
Best Local Similarity 26.7%; Pred. No. 4e-24;
Matches 130; Conservative 72; Mismatches 234; Indels 50; Gaps 13;

QY 4 PHVLTTPAQGHINPALQAKNL-----VKMGIEVTSTSYAOSRMDEKSIILNA----P 55
D 12 PHAVAVAFPESSHAAYLLSTARALAAAAAPGATLSFLSTASSLAQLKASSASAGHGLP 71
QY 56 KGLNFIPIFSGDEGDEHSDKDPVYNSQLRKCGSET--VKKII--LTCSENGQIPITCLLY 111
D 72 GNLFREVEVPDGPAA--EETVPVPRQMLFMEAAEAGGVKAWLEAARAAAGARVTCVG 129
QY 112 SIPLUPAAAEVAREVHPSALLWSOPATILDIIYFNPHGYEKAMANESND-----PNWSI 165
D 130 DAFVWPADAASAGAPVVPVWTAASCAL-----LAHRTDALREDVGDQAAANRVGILLI 184
QY 166 QLPGLPLETRDLPSPFLPYGAKGSLRVALPFPKELIDTDAETTPKTLVNTFDELEPEA 225
D 185 SHPLASRYVRDLDPGVV---SGDFNYVINLLVHRMGQCLPRSAVAALNTFPGLDPPD 240
QY 226 LNATEGYKFGIGPLIPSAFLGNDPLDASFGDGLFQNSNDYMEWLNKPNSSVYVIFSFG 285
D 241 VTAALAEILPNCVFPFGPYHLLAEDDADTAAPAD-----PHGCLAWLGHQPARGVAYYSFG 296
QY 286 SLMNPISQMEIEISKGLIDIGRPFLLWIKENE-----KGKEENKKGICIELEKIGKIY 340
D 297 TVACPRDELELRELAAGLEDSGAPFLWSLREDSWPHLPFGFLDRAAGTGS-----GLVV 349
QY 341 PWCSOLEVLKPSLCGVSHCGNWSALESACGVVPVAFPPQWTDQMTNAKQVDEWVKSQV 400
D 350 PWAPQVAVLHPSPGVAFVHAGWASVLEGLSGVPACPRPFEGDQRMNARSVAHVNGFEGA 409
QY 401 RVRINEDGVVSEIEIKRCIELVMDGGEELKNAKWKELAREAVKEGSSSHKLNKAF 460
D 410 AF-----ECAMTSAGVATAVEELL-RGEGARMRAKAKELQALVAEAFGPGGECRKNFDRF 464
QY 461 IDDVAK 466
D 465 VEIVCR 470

RESULT 8
UFOL_MAIZE STANDARD; PRT; 471 AA.
ID UFOL_MAIZE AC P16167;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE 3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-W22 ALLELE).
OS BZ1 OR UGT71A1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88284304.
RA Ralston E.J., English J.J., Dooner H.K.;
RT "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL Plant Mol. Biol. 11:473-481(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88284304.
RA Ralston E.J., English J.J., Dooner H.K.;

```


[illegible][illegible]


```
Db 72 GNLREVEVPGAPAA--EETVPVPRQMLFMEAAEAGGVKAWLEAARAAAGARVTCVVG 129
QY 112 SIFLPMAAEVAREVHIPSALLWSQPATILDIYFNFHGYEKAM-----ANESNDPNWSIQL 167
Db 130 DAFVMPADAASAAGVWPVWTAASAL-LAHIRTDLSREDVGQAAANRDEP--LISH 186
QY 168 PGLPLETRDLPFLPYGAKGSLRVALPFPKELIDTLDATTPPKILVNTFDELEPEALN 227
Db 187 PGLASRYRDLDPGVV-----SGDFNVISLLVHMGCCLPRSAVAALNTPGLDPPDVT 242
QY 228 AIEGKYFGIGLIPSLAFLGGNDLPDASGGDLFQNSNDYMENLNSKPNSSVYISFSL 287
Db 243 AALAEILPNCVPFPGPYHLLAEDDADTAAPAD-----PHGLAWLGRQARGVAVYSGTV 298
QY 288 MNPSTISOMERISKGLIDIGRPFELWTKENE-----KGKEENKKGCTEELKIGKIVPW 342
Db 299 ACPRDELRELAALEASAPFLWSLREDSWLLPPGFLDRAAGTGS-----GLVFPW 351
QY 343 CSQLEVLKHPSLGCFVSHCGNSGALSGLACGVVPVAFPOWTDQMTNAKQVEDVWKSGVRV 402
Db 352 APOAVLHPHPSGAFVTHAGWASVYEGYSGVPMACRPFPGDQRMNARSVAHYVGFGAAF 411
QY 403 RINEDGVVSEIEIKRCITELVMDGGEELRNKAKWKELAREAVKEGGSSHKNLKAFID 462
Db 412 ---EGAMTSAGVAAVEELL-RGEGAGHRARAKELQALVAEATPGGECRKNFDEVE 466
QY 463 DVAK 466
Db 467 IVCR 470

RESULT 12
UFO6_MANES
ID UFO7_MANES STANDARD; PRT; 394 AA.
AC Q40288;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 6 (EC 2.4.1.91) (UDP-GLUCOSE
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 6) (FRAGMENT).
GN GT6 OR UGT73A6.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
OC Euphorbiaceae; Manihot.
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=COTYLEDON;
RX MEDLINE; 95201291.
RA Hughes J., Hughes M.A.;
RT *Multiple secondary plant product UDP-glucose glucosyltransferase
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.*;
RL DNA Seq. 5:41-49(1994).
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC INSTABLE CYANOHYDRINS TO PRODUCE STABLE CYANOGUCOSIDES.
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS AND LEAVES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN DEVELOPING COTYLEDONS
CC (FROM EMERGING GREEN COTYLEDONS TO APPROXIMATELY 10 DAYS OLD);
CC VERY LOW LEVELS IN HYPOCOTYLS AND NO MEASURABLE EXPRESSION IN
CC ROOTS.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77463; CAAS4613.1; -.
DR PFAM; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Multigene family.
FT NON_TER 1
SQ SEQUENCE 394 AA; 44264 MW; FDSICD974E5ADCD7 CRC64;

Query Match 15.9%; Score 393.5; DB 1; Length 394;
Best Local Similarity 29.6%; Pred. No. 1.8e-22;
Matches 118; Conservative 76; Mismatches 155; Indels 49; Gaps 14;

QY 90 ETVKKIILITCSENGQPTCLLYSTFLPMAAEVAREVHIPSALLWSQPATILDIYVF--NP 147
Db 18 EAVSKL---TARSSSLAGFVLDMFCTSMIDVAKELGVPYIYFTTSGAAFLGFLFYVOLI 74
QY 148 HGYEKAMANESNDPNWSIQLPGLPLETRDLPFLPYGAKGSLRVALPFPKELIDTLD 207
Db 75 HDEQDAULTQFKDSABLSVPSL-----ANSLPARVLP--ASMLVKDRFYAFIRIRGL-- 126
QY 208 ETPKILVNTFDELEPEALNAIEGK-----FYGIGLIPSLAFLGGNDLPDASFGDLFQ 262
Db 127 REAKGIWNTFPELESALNSLKDQSKIPPYPVGPILK---LSNQE-----NDVGP 176
QY 263 NSNDYMEWLNKPNSSVYISFGSLMNPISQMEIEISKGLIDIGRPFELWTKENE-NEKKE 321
Db 177 EGSEIIEWLDQPPSSVFLCFGSMGFMQDAKEACALEQSRHRLMSLRPPPKGK1 236
QY 322 E-----ENKK-----LGCIEELEKIGKIVPWCSOLEVLKHPSLGCFVSHCGNSALES 372
Db 237 ERTSTDYENLQELLPYGFSERTAGKGVVGHAPQVAILEHPAIGGVSHCGNSILESIFW 296
QY 373 GVPVVAFPQWTDQMTNAKQVEDVWKSGVRVINEGDGVESE-----ETKRCIELVMDGG 426
Db 297 SVPTATWPLVYAEQOOFNFTM--VTGLGLAVEIKMKYKESEIILSADDIERGKCMV--- 351
QY 427 EKGEELRNKAKWKELAREAVKEGGSSHKNLKAFIDV 464
Db 352 EHSIRKRVKEMSKRKALMDDESSFWLDRLEDV 389

RESULT 13
UFO7_MANES
ID UFO7_MANES STANDARD; PRT; 287 AA.
AC Q40289;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 7 (EC 2.4.1.91) (UDP-GLUCOSE
DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 7) (FRAGMENT).
GN GT7 OR UGT73A7.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Malpighiales;
OC Euphorbiaceae; Manihot.
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=COTYLEDON;
RX MEDLINE; 95201291.
RA Hughes J., Hughes M.A.;
RT *Multiple secondary plant product UDP-glucose glucosyltransferase
RT genes expressed in cassava (Manihot esculenta Crantz) cotyledons.*;
RL DNA Seq. 5:41-49(1994).
CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC PIGMENTS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL
CC 3-O-D-GLUCOSIDE.
CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
```


OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
 CC Solanaceae; Petunia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, V-26; TISSUE=LEAF;

RX MEDLINE: 94177177.

RA Kroon J., Souer E., de Graaff A., Xue Y., Mol J., Koes R.;

RT "Cloning and structural analysis of the anthocyanin pigmentation
 locus Rt of Petunia hybrida: characterization of insertion sequences
 in two mutant alleles.";
 RL Plant J. 5:69-80(1994).
 RN [2]

RP SEQUENCE OF 3-473 FROM N.A.

RC STRAIN=CV, OLD GLORY BLUE; TISSUE=PETAL;

RX MEDLINE: 94177178.

RA Brugliera F., Holton T.A., Stevenson T.W., Farcy E., Lu C.Y.,

RA Cornish E.C.;
 RT "Isolation and characterization of a cDNA clone corresponding to the
 Rt locus of Petunia hybrida.";
 RL Plant J. 5:81-92(1994).

CC -1- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
 GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
 PIGMENTS.

CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL -> UDP + FLAVONOL

CC -1- 3'-O-D-GLUCOSIDE.

CC -1- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE

CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.

CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X71060; CAA50377.1; -

DR EMBL; X71059; CAA50376.1; -

DR EMBL; Z25802; CAA81057.1; -

DR PFAM; PF00201; UDPGT; 1.

DR PROSITE; PS00375; UDPGT; 1.

RW Transferase; Glycosyltransferase.

FT CONFLICT 246 255 LIGPVVDPDP -> SNRTRSSGPA (IN REF. 1).

FT CONFLICT 348 348 N -> H (IN REF. 2).

FT CONFLICT 431 431 D -> E (IN REF. 2).

SQ SEQUENCE 473 AA; 52405 MW; A310E4BFBC9EB02A CRC64;

Query Match 15.28; Score 375; DB 1; Length 473;
 Best Local Similarity 28.08; Pred. No. 5.9e-21;
 Matches 130; Conservative 75; Mismatches 200; Indels 60; Gaps 16;

QY 5 HVIITPTPAQGHINPALQFAKNLVKMGIEVTESTIYAQSRMDEKSTILNAPKGLNFIPTFS 64

Db 13 HVVMPFPFAFGHISPFVQLANKLSSYGVKVSFTASGNASRV--KSLNLSAPTTTHIVPLT 70

QY 65 ----DGEDEGHDSDKDPVFMNSQLRKCSETVKKIILTCSENGQITCLLYSIFLPWAAE 120

Db 71 LPHVEGLPPGAESTAEITPASAELLVALDLMOQIKTLHLKRP-HEVLFDFAGQEWLPK 129

QY 121 VAREVHIPISALLMSQPATILDIYFFNFGHYEKAMANESNDPNWSTQLPGLPLETRDLPS 180

Db 130 MANGLGIKT-----VYYSVVVALSTAFL--TCPARVLEPKKYPSPLEDMKKPP 174

QY 181 FLIPYGAAGSLRV-----ALPPFKE-----LIDTLDA--ETTPKILVNTFDELEPALN 227

Db 175 LGFPQTSVTSVTFTEARDFLYVFKSFHNGPTLYDRIQGLRGCSAILAKTCSQMEGPYIK 234

QY 228 AIEGY----KFYIGIGLIPSAFLGNDPLDASFGGLDFONSDYMEWLNKSPNSVWYISF 284

Db 235 VYEAQFNKPVFLIGPVVP-----DPPSGKL-----EEKWATWLNKFEGGTVIYCSF 280

QY 285 GSLMNPISQMEIEISKGLDIDGRPFLMWIK--ENEKGKEENKKL--CCIEELEKIKIKI- 339

Db 281 GSETFLTDOQVKELALGLEQTGLPFFLVLPANVDVSAELNRLALPEGLFRLVRDKGIIH 340

QY 340 VPWCSQLEVLKHPISLGCFSVSHCGWSALESACGVPVVAFPQWTDQMTNAKQVEDVWVKS 399

Db 341 SGWVQQQNILLAHSSVCGYVCHAGFSVIEALYNDQCQVYMLPKQGDQILNAKLVSMDMEAG 400

QY 400 VRV-RINEDGVVSESEIKRCIELVMDGGEK--GEELRNKNAKWKKE 441

Db 401 VEINRDEDSYFGKEDIKAVEKVMYVDVEKDFGKLIENQKKWKE 445

Search completed: August 1, 2000, 21:32:28

Job time: 6048 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2000, 14:38:44 ; Search time 3616.02 Seconds
(without alignments)
2567.101 Million cell updates/sec

Title: US-09-147-955-11
Perfect score: 2105
Sequence: 1 agtgcgcgaacgcaattaa.....acaaaaaaaaaaaaaaaaa 2105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

[illegible]


```

/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="X11-Blue MRP"
/note="vector: pBlueScript SK(-); Site.1: EcoRI; Site.2:
XhoI; cLED - Tomato Carpel EST Library. Oligodm-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
196 a 82 c 114 g 154 t
BASE COUNT

```

```

Query Match      15.4%; Score 325; DB 45; Length 546;
Best Local Similarity 79.2%; Pred. No. 2.7e-66;
Matches 414; Conservative 0; Mismatches 100; Indels 9; Gaps 2;

Qy 913 tcgagtgctcactccaccattcaagaagttagagacacattagatgctgaaaccactcc 972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30 TAGTTTTGCTCTACCAACATTCAAAGAGCAACTAGACACATTAGATGGTGAAGAAATCC 89

Qy 973 taagatctcttgatatacaatttgatgaattagagcctgagcactcaatgcaattgaagg 1032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 AAAGGTACTGTGGAATACATTTGATGTGCATTAGACTAGAGCCACTCAAGACTATTGCAAAA 149

Qy 1033 ttataagttttatggaattggaccggtgattccctctgcttcttgggtggaaaagaccc 1092
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 ATACAATTTAATTGGAATTTGGACCATGATTTCCTTCATCATCTTGGTGGAAAGATTC 209

Qy 1093 tttagatgcttcatttgggtggtgattcttttcaaaatctaa---atgactatatggaatg 1149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 ATTGGAATCTTCATTGGTGGTGATCTTTTTCAAAAGTCAAAATGATGACTACATGGAATG 269

Qy 1150 gtcaaaactcaaaagcctcaatcatcagttgtttatatattcttgggagctcaatgaatcc 1209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 GTTAAACACAAAGCCTAAATCATCAATTCATTGCTTATATCTCATTTGGGAGTCTATTCAATTT 329

Qy 1210 atctattagccaaatggaggagagatatcnaaagggttagatagacatatgaaggccgttttt 1269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 ATCAGAAACCCAAAAGGAGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATTCTT 389

Qy 1270 atgggttgataaaagaaaaatgaaaaggcnaaagagaagagaaataaaaagcttggttgat 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 ATSGGTAAATAAGAGATCCGAAAGAAGAAAAAGAAAGAGAG-----AAATTCAAGTTGCAT 443

Qy 1330 tgaagaaatggaaaaaaataggaaaaaatagttccctgggtctcaaacacttgaagttctaaa 1389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 GATGGAAATTAGAGAGCAAGGAAATAATAGTACCATTGTGTTCAACAATTTGAAAGTCTGCAC 503

Qy 1390 acatccacttttagatgatttttctctcatgttggaatggaat 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 ACATCCATCTTTAGGATGTCGTGCTCGCACTGTGGATGGAAT 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

Unpublished (2000)
 On Jan 6, 2000 this sequence version replaced gl:5677182.
 Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU
 5 prime sequence.

FEATURES
 source
 Location/Qualifiers
 1..658
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLE116K15"
 /clone_lib="tomato germinating seedlings, TAMU"
 /tissue_type="whole seedlings"
 /dev_stage="7 days post imblition"
 /note="vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
 XhoI; 7 days post imblition on water-agar. Mixed stage
 whole germinating seedlings from seed coat emergence up
 to two centimeters in seeds not showing obvious signs of
 germination were discarded."

BASE COUNT 208 a 126 c 124 g 200 t
 ORIGIN

Query Match 15.0%; Score 315.4; DB 79; Length 658;
 Best Local Similarity 71.7%; Pred. No. 5.2e-64;
 Matches 471; Conservative 0; Mismatches 176; Indels 10; Gaps 4;

QY 566 gaccctgtatttacatgcacaaactcttgtaaatgtgaagtgaatacttcacaaaaata 625
 Db 2 GATCTAAACGTTACATGTCAGAGATGAAGAAGTCGTGGTCCCAACATGAGGGATGTC 61
 QY 626 attctcaactgtctgaaaaaggacagcgcctataaactgctactttactccatttcctt 685
 Db 62 ATTTGAAGAGTTCAGACGAGGAGCGTCTCTGCACGTCCTCGTCTACACTCTCTGCTTA 121
 QY 686 ccttggcgcagaggttagcagtggaagttcacactcctctcctctctctctcttcttggagtc 745
 Db 122 CCTTGGCGACGTGAGTACGGGTGAATCCATTCCTCCATCCCGGTCTATGAATTCAA 181
 QY 746 ccagcacacaattgggacatatattcacttcaacttcttcattggatgataaaagctat--g 802
 Db 182 CCAGCAACTGTGCTAGACATATACATCTATTTACTTCAATGGCTATGAGGATGAATGAAG 241
 QY 803 gctaataatccaatgataccaaattggctccactcaactcccgggcttcacactgcgaa 862
 Db 242 TGTACTACCTCAANTGATCCAAATTTGGAGTATCCCAATGCGAAGGTTTCATCTACTATA 301
 QY 863 actcgagatcttctctc-attttactccttattggtgcaaaagggagt---cttcgagt 918
 Db 302 AGCCAAGATCTTCCATCTTTTTTTAGTTTCACTATAGCTCAAAAAGATGATGATAGTTT 361
 QY 919 tgcactccacattcaagaattgatagacacattagatgctgaaccactcccaagat 978
 Db 362 TGCTCTACCAACATTCAGAGCAACTAGACACATTAGATTGGTGGAGAAATCCCAANGGT 421
 QY 979 tcttgtgaatacatttgatgaattagagccttgaggcactcaatgcgaattgaagggttataa 1038
 Db 422 ACTTGTCATATATTGATGCAATTAGACTAGAGCCACTCAAAAGCTATTGAAAAATACAA 481
 QY 1039 gttttatggaattgacccgtgtaccctctctgcttcttgggtggaaatgaccccttcaga 1098
 Db 482 TTTAATTTGGAAATGGACCATGATTCCTTCATCATCTCTGGGTGGAAAGATTCATTGGA 541
 QY 1099 tgcctcatttgggtgatctcttttcaaaatccaa---atgcactatgatgaattggttaa 1155
 Db 542 ATCTTCATTGTGGTGTGATCTTTTTCAAAAGTCAAAATGATGACTAGATGAATGTGTTAA 601
 QY 1156 ctcaagccaaattccatcagttgtttatatatatcttttggagcgtcctaatgaaaccatc 1212

Db	277	GTTAAACCAAGCCTAATATCATCAATTGTTTTATATATATATTCATTTGGGAGTCTATTGAATTT	336
QY	1210	atctattgccaataagagagagatatcaaaagggttgatagacataggaagccgtttt	1269
Db	337	ATCAAGAACAACCAAGAGAGAGATTCCAAAAGGGTGATAGATATCCAAGGCCATCTT	396
QY	1270	atgggtgatataagaataatgaaaaggcacaagaagaagaataaaaaagcttggttat	1329
Db	397	ATGGGTATATAGATATCAAGACAGAAAAAGACAGACAGAAATTA-----GTTCAT	450
QY	1330	tgaagaattggaataaataggaaaaatagtcttcattggttcacaacttgaagttctaaa	1399
Db	451	GATGGAATTAGAGAAGCAAGGAAAAATAGTACCATGGTGTTCACAACTTGAAGTCTGCAC	510
QY	1390	acatccatct 1399 	
Db	511	ACATCCATCT 520 	
RESULT	9		
AI899102			
LOCUS	AI899102	478 bp mRNA EST	27-JUL-1999
DEFINITION	EST268545 tomato ovary, TAMU Lycopersicon esculentum cDNA clone.		
ACCESSION	CLD37A16, mRNA sequence.		
VERSION	AI899102		
KEYWORDS	AI899102.1 GI:5605004		
SOURCE	EST.		
ORGANISM	tomato.		
REFERENCE	Lycopersicon esculentum		
AUTHORS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 478)		
TITLE	Alcala J., Vrebalov J., White R., Matern A.L., Vision T., Holt L.E., Liang P., Upton J., Ronning C.M., Craven M.B., Fujii C.Y., Bowman C.L., Nierman W., Fraser C.M., Venter J.C., Martin G.B., Tanksley S.D. and Giovannoni J. Generation of ESTs from tomato carpel tissue Unpublished (1999) On Feb 18, 1999 this sequence version replaced gi:4296853.		
JOURNAL	Contact: David Frisch		
COMMENT	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU 5 prime sequence.		
FEATURES	Location/Qualifiers		
source	1..478		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="CLD37A16"		
	/tissue_type="carpel"		
	/dev_stage="5 days pre-anthesis to 5 days post-anthesis"		
	/lab_host="Xil-Blue MRF"		
	/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site xhoI: cLED - Tomato Carpel EST Library. Oligodm-primed directionally cloned cDNA in vector Lambda ZAP II with and 3' ends located at the EcoRI and XhoI sites, respectively."		
BASE COUNT	172 a 67 c 102 g 137 t		
ORIGIN			
Query Match	14.3%; Score 301.4; DB 45; Length 478;		
Best Local Similarity	80.1%; Pred. No. 9.5e-61;		
Matches 382; Conservative	0; Mismatches 86; Indels 9; Gaps		
QY	961	tgaaccaccttaagattcttgatacatattgaaattagagctgaggacactcaa	1020

```

Db 2 TGAAGAAATCAAAGGCTACTTGTGAATACATTTGATGATCATTTAGAGCTAGAGCCACTCAA 61
QY 1021 tgaattgaaggttataagttttatggaaattgagccggttgatcctctgtctttctggg 1080
Db 62 AGCTATTGAAAATACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 121
QY 1081 tgaataagaccttagatgcttcatcttggtggtgatcttttcaaaattcaa---atga 1137
Db 122 TGAAGAAATCATTCGAAATCTTCATTTTGGTGGTATCTTTTCAAAAGCTCAAAATGATGA 181
QY 1138 ctatatgaattggttaactcaaaagccaaattcatcagttgttttatatatcttttggag 1197
Db 182 CTACATGAATGGTAAACACAAAGCCTAATCATCAATTTGTTTATATCTCATTTGGGAG 241
QY 1198 tctaataatccatctattagccaaattgaggagatcatcaaaaggttgatagacatagg 1257
Db 242 TCTATTGAATTTATCAAGAAACCAAGAGGAGAGATGTCAAAAGGGTGTATAGATCCA 301
QY 1258 aagccgctttttatgggtgataaagaataatgaaaagcgaaggaaggaagataaaaa 1317
Db 302 AAGGCCATCTTCTTGGGTAAATAGAGATCAAGAAAGAAAGAAAGAAAGAAATTA-- 360
QY 1318 gcttggtgtattgaagaattgaaaaataggaaaaatagttccatggtgttcacaa 1377
Db 361 -----GTTGCGATGTGGAATAGAGAGCAAGGGAATAGTACCATGGTGTTCACAACT 415
QY 1378 tgaagttcaaaacatccatttttaggtatgttttggtttcttcttcttcttcttcttctt 1434
Db 416 TGAAGCTCTGACATCCATCTTTAAGATGTTTTTGTCTGCACCTGTGGATTC 472

RESULT 10
LOCUS A1779099 517 bp mRNA EST 29-JUN-1999
DEFINITION EST259978 tomato susceptible, Cornell Lycopersicon esculentum cDNA
clone cLES719, mRNA sequence.
ACCESSION A1779099
VERSION A1779099.1 GI:5277140
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (Bases 1 to 517)
D' Ascenzo, M., He, X., Lyman, J., Matern, A. L., Vision, T., Hoit, I. E.,
Liang, F., Upton, J., Ronning, C. M., Craven, M. B., Fujii, C. Y.,
Bowman, C. L., Nierman, W., Fraser, C. M., Venter, J. C., Tanksley, S. D.,
Giovannoni, J. J. and Martin, G. B.
Generation of ESTs from Pseudomonas susceptible tomato
Unpublished (1999)
On Jan 19, 1998 this sequence version replaced gi:2285531.
Contact: David Frisch
Clemson University
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
Location/Qualifiers
1..517
/organism="Lycopersicon esculentum"
/cultivar="R11-13 (Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="cLES719"
/clone_lib="tomato susceptible, Cornell"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLES - Tomato Pseudomonas Susceptible EST Library.
Directionally cloned cDNAs inserted into pBluescript

```

```

BASE COUNT SK(·) at 5' end with EcoRI and 3' end with XhoI site*
ORIGIN 185 a 78 c 102 g 152 t
Query Match 14.0%; Score 294.6; DB 44; Length 517;
Best Local Similarity 75.8%; Pred. No. 3.8e-59;
Matches 392; Conservative 0; Mismatches 119; Indels 6; Gaps 2;
QY 785 ggaatgaaaaagctatggttaataagatccaatgataccaaattggtccaatccaattcc 844
Db 1 GGTGAGGATGAATGAAGTGTAGTGTAGCTCAATTAATGATCCAAATGGAGTATCCAAATGGCA 60
QY 845 gggcttccactactggaaactcgagatcttcccttcttcttcttcttcttcttcttcttct 904
Db 61 AGGCTTCCATTTACTTAAAGCAAGATCTTCCATCTTTTGTAGTTTCATCTAGCTCAAAA 120
QY 905 ---ggaggtcttcaggttgcacttccaccattccaagaattgatgacacatagatgct 961
Db 121 GATGATAAGTATAGTTTGTCTTACCAACATTCAAAGAGCACTAGACACATTAGATGCT 180
QY 962 gaacacactcctaagattcttggatacacatttgatgaattagagcgtgagcactcaat 1021
Db 181 GAAGAAATCCAAAGGTACTTGTCAATACATTTGATGATGATGATGATGATGATGATGAT 240
QY 1022 gcaatggaaggttataagtttttggaaattggaccgttgatcttcttcttcttcttctt 1081
Db 241 GCTATTGAAAAATCAATTTAAATTTGAAATTTGGACCATTTGATTCCTTCATCATCTTGGGT 300
QY 1082 ggaactgaccttttagatgcttcttcttcttcttcttcttcttcttcttcttcttcttct 1138
Db 301 GCAAAAGATTCATTGGGAATCTTCATTTGGTGGTGAATCTTTTCAAAAGTCAAAATGATGAC 360
QY 1139 tatatggaatgggttaactcaaaccaaaccaaaccaaaccaaaccaaaccaaaccaaacca 1198
Db 361 TACATGGAATGGTTAAACACAAAGCCCTAAATCATCAATTTGTTTATATCATCTTGGGAGT 420
QY 1199 ctaatggaatccatctattagcgaatggaggagatatacaaaagggttgatagacatagg 1258
Db 421 CTATTGAATTTATCAAGAAACCAAAAGGAGAGATTCAAAAGGGTTGTAGATGATCCAA 480
QY 1259 aggcggttttatgggtgataaagaataaagaataaagaataaagaataaagaataaaga 1295
Db 481 AGGCCATCTTTATGGGTAAAGAGATCAAGAGAAG 517

RESULT 11
LOCUS A1489014 500 bp mRNA EST 29-JUN-1999
DEFINITION EST247353 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
cLED19F7, mRNA sequence.
ACCESSION A1489014
VERSION A1489014.1 GI:4384385
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanales; Solanum; Lycopersicon.
1 (Bases 1 to 500)
Alcala, J., Vrebalov, J., White, R., Matern, A. L., Vision, T.,
Holt, I. E., Liang, F., Upton, J., Ronning, C. M., Craven, M. B.,
Fujii, C. Y., Bowman, C. L., Nierman, W., Fraser, C. M., Venter, J. C.,
Martin, G. B., Tanksley, S. D. and Giovannoni, J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
On Nov 2, 1998 this sequence version replaced gi:3630982.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293

```

Email: dfrisch@CLEMSON.EDU

	FEATURES	SOURCE
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
35.
36.
37.
38.
39.
40.
41.
42.
43.
44.
45.
46.
47.
48.
49.
50.
51.
52.
53.
54.
55.
56.
57.
58.
59.
60.
61.
62.
63.
64.
65.
66.
67.
68.
69.
70.
71.
72.
73.
74.
75.
76.
77.
78.
79.
80.
81.
82.
83.
84.
85.
86.
87.
88.
89.
90.
91.
92.
93.
94.
95.
96.
97.
98.
99.
100.

```

FEATURES                                source
1. .500
   Location/Qualifiers
      /organism="Lycopersicon esculentum"
      /cultivar="TA496"
      /db_xref="taxon:4081"
      /clone="cLED19F7"
      /clone_lib="tomato ovary, TAMU"
      /tissue_type="carpel"
      /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
      /lab_host="X11-Blue MRF"
      /note="vector: pBlueScript SK(-); Site_1: EcoRI, Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5',
and 3' ends located at the EcoRI and XhoI sites,"

```

Query match 13.6%; Score 286.4; DB 40; Length 500;
Best Local Similarity 78.7%; Pred. No. 3.2e-57;
Matches 370; Conservative 0; Mismatches 91; Indels 9; Gaps 2;

913	QY	tcgagttgcacttcacacattcaagaattgatatagacacattagatcgtgaacccactcc	972
37	Db	TAGTTTTGGCTTACCACCAATTCAAAGAGCAACTATAGACACATTAGATTGGTGAAGAANAATCC	96
973	QY	taagattctctgtaatacatcttgatgaattagagcctgagggcactcaatgcaatggaagg	1032
97	Db	AAAGGTACTTGTGAATACATTTTATGCAATTTAGAGCTTAGAGCCCACTCAAAGACTATTTGAAAA	156
1033	QY	ttataagttttatggaaatggaccgcttgattcctctctggtttgggtggaataagccc	1092
157	Db	ATACAAATTTAAATTGGAATTTGGACCACTGATTCCTTCATCATTTCTGGGTGGAAAAAGATTTC	216
1093	QY	tttaqetgcttcatttggtggtgatctttttcaaaattcaa---atgacatatatgaatg	1149
217	Db	ATTGGAAATCTTCATTTGGTGGTGATCTTTTCAAAAGTCAAATGATGATCTACATGGAAATG	276
1150	QY	gttaaaactcaaaagcacaattcatcagttgtttatatatatctttgggagctcaatgaatcc	1209
277	Db	GTTAAACACAAAGCCTAAATCATCAATTTGTTTATATCTCATTTGGGAGTCTATTGAATTT	336
1210	QY	atctattagcacaattgaggagagatatcaaaaggttgatagacataggagccggttttt	1269
337	Db	ATCAAGAAACCAAAAGGAGGAGATGTCAAAGGGTGTATAGATCCAAAGGCCCACTTCCT	396
1270	QY	atgggttgataaagaataatgaaagcgaagaagaagagaaataaaagcttggttgat	1329
397	Db	ATGGGTATATAGAGATCAAGAAAGAGAAAAAGAAAGAGAGAAAAATTAA-----GTTGCAT	450
1330	QY	tgaagattggaaaaataaggaataatggttcctatggtgttcacaaacttg	1379
451	Db	GATGGAAATTAGAAGACGAAGGAAAAATAGTACCATGTTGTTTCACAACTTG	500

RESULT	12
AI483541	
LOCUS	476 bp mRNA
DEFINITION	AI483541 tomato ovary, TAMU EST 29-JUN-1999
ACCESSION	U024390 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
VERSION	CLD25M13, mRNA sequence.
KEYWORDS	AI483541
	AI483541.1 GI:4387465
	EST.

Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C.,
Marlin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato carpel tissue
unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@clemson.edu

Query Match 12.5%; Score 263; DB 40; Length 476;
Best Local Similarity 79.6%; Pred. No. 1e-51;
Matches 324; Conservative 0; Mismatches 80; Indels 3; Gaps 1;

Qy	913	tcgagttgcacttccaccattccaagaattgtagacacattagatcgtgaaacacattcc	972
Db	37	TAGTTTGGCTTACCACGATTCAAAGAGCAACTAGACACATTAGATTGGTGAAGAAATCC	96
Qy	973	taagatctcttgaaatacatttgatgaattagagcctgaggcactcaatgcaaatgaag	1032
Db	97	AAGGTACTTGTGAATACATTTGATGCATTTAGAGCTTAGAGCCACTCAAGACTATTGAAAA	156
Qy	1033	ttataagttttatggaattgacccgttgattccctctgcttctcttgggtggaaatgaccc	1092
Db	157	ATACAATTTAATTGGAATTGGACCACTGATTCCTTCATCATCTCTGGGTGGAAAGATTTC	216
Qy	1093	tttagatccttaattgggtggatctcttttccaaatccaa---atgactatatggaatg	1149
Db	217	ATTGGAATCTTCATTTGGTGTGATCTTTTTCAAAAGTCAAAATGATGACTACATGGAAATG	276
Qy	1150	gttaaacctcaagccaaattcatcagttttatatatcttttggagctcctaagaaacc	1209
Db	277	GTTAAACACAAAGCCTAATCATCATTTGTTTATATCTCATTTGGAGGTCATTGGAATTT	336
Qy	1210	atctattagccaaatggaggagatatacaaaagggtttgatagacataggaaagccggtttt	1269
Db	337	ATCAAGAAACCAAAAGGAGGAGATTGCAAAAGGGTTGATAGAGATCCAAAGGCCATCTT	396
Qy	1270	atgggttgataaaagaaatgaaaaggccaagaagaagagataanaa	1316
Db	397	ATGGGTGAATTAGAGATCAAGAAGAAGAAAAGAAAGAAAGAGAATAATTA	443

RESULT	13
AW349414/c	
LOCUS	747 bp mRNA EST O1-FEB-2000
DEFINITION	GM210007A20D2R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
ACCESSION	AW349414
VERSION	AW349414.1 GI:5847124
KEYWORDS	EST.
SOURCE	soybean.
ORGANISM	Glycine max
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
 Fabaceae; Papilionoideae; Glycine.

REFERENCE

1 (bases 1 to 747)
 Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
 Erpelnding, J., Raph, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)

JOURNAL

On Jul 9, 1999 this sequence version replaced gi:5434916.
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)

COMMENT

Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics

University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4533 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

Location/Qualifiers
 1..747
 /organism="Glycine max"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone_lib="Gm-r1021"
 /tissue_type="root"
 /lab_host="XL10-Gold"

/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
 XhoI; Library Gm-r1021 is a sequence-driven, roracked set
 of the original library Gm-cl004 which was prepared from
 day old 'Williams' seedlings which were propagated on
 paper towels with distilled water. Stratagene's cDNA
 Synthesis Kit (catalog #200401) was used to synthesize
 the cDNA. The Gm-cl004 library was constructed by Dr.
 Paul Keim & Virginia H. Coryell, Department of Biology,
 Box 5640, Northern Arizona University, Flagstaff, AZ
 86011, email: paul.keim@naui.edu, virginia.coryell@naui.edu.
 The contig analysis to select unique genes was performed
 by the laboratory of Ernest Retzel, Computational Biology
 Centers, University of Minnesota,
 http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
 . Reracking was performed by Genome Systems, St. Louis,
 http://www.genomesystems.com, and sequencing by the Keck
 Center for Comparative and Functional Genomics,
 University of Illinois,
 http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT

206 a 235 c 97 g 187 t 22 others

ORIGIN

Query Match 11.8%; Score 247.4; DB 71; Length 747;
 Best Local Similarity 61.7%; Pred. No. 5.2e-48;
 Matches 424; Conservative 0; Mismatches 236; Indels 27; Gaps 2;
 QY 1061 attccttctgtttcttgggtggaatgaccttttagatgcttcttgggtggtattt 1120
 DB 733 ATTCTNNNGNCNTTNCNNNGCGGTAAAGATCCCGCGGACTACTNNNTATGCGGTGATNNN 674
 QY 1121 tttcaaaattcaatgactatagatgggtttaaactcaaacgcaattcatcagttgtt 1180
 DB 673 NNCNNTGCTCAATATGATACGTTGAATGGTTGGACTACACGCTGAGTTATCTGTGTTT 614
 QY 1181 tatatatcttttggaggtctaataatccatctattagccaatggagagatatcaaaa 1240
 DB 613 TATGTTTCATTTTGTACCCCTGCTGTGTGTGGCTGATAGACAGATGAAGAGAGCTTGCACGC 554
 QY 1241 ggggttgatagacataggaagccgctttttatgggtgataaaagaaaaagaaagc 1300

Db 553 GGCCTGCTAGATTCGGGATATCTCTTCTTGTGGGTCAATTAGAGATATGCAAGAAATAGAA 494
 QY 1301 gaagaagaataaaaaagctgggtgtattgaagaatttgaaaaaatagaaaaatagtt 1360
 Db 493 GA-----TAAGTCAGAGAGAACTGGAGCAGAGGGGTAAAGATTGTG 452
 QY 1361 caatggtttcacaaactgaagttcttaaaacatccatcttttaggatgtttttcttc 1420
 Db 451 AAATGGTGTCTCAGTGGAGGATTCCTGTCATGGTTCCTTGGGTGTTTGTAAACGAT 392
 QY 1421 tgggatggaattcagccttagagagttttagctgttggagtgccagtttggcatttct 1480
 Db 391 TGTGATGGAATTCGACTATGGAAGTTTGGTTCGGGGTTCCTATGTCGCTTCCG 332
 QY 1481 caatgacagatcaaatgcaaaatgccaacaagttgaagaatgtggaagaatggaata 1540
 Db 331 CAGTCACACACACACAGGACGATGCAAAAGATGTCACAGATGTGTGGAGAGGGGGTG 272
 QY 1541 agagtgaagaataaa-----tgaagatgggtgttgaagaatgaggaatcaaaagg 1591
 Db 271 AGGTTGGATGATAAGGTCAATGTTGAGGAAGGATAGTGAAGCAGAGGAGATTAGGAAG 212
 QY 1592 tgaattgaattggaatggaatggaagagagaggaagaagaattgagaagaatgactaag 1651
 Db 211 TGTGTTGGATGTGTCATGGGGAGTGGAGGAAAGACAGGAATTCAGAGGAATGCTGAT 152
 QY 1652 aaatggaagaattggctagagaagctgtggaagaagggtggtctcttcacacaagaattta 1711
 Db 151 AAATGGAATGTTGGCCAGGAAAGCGTCACGGAAGAGGCTCTTCGGATAGTAATATG 92
 QY 1712 aaggtttttattgatgtgttggccaaa 1738
 Db 91 AGGACTTTTCTCCATGATGTGTCANAA 65
 RESULT 14
 AW030131
 LOCUS AW030131 650 bp mRNA EST 15-SEP-1999
 DEFINITION ES273386 tomato callus, TAMU Lycopersicon esculentum CDNA clone
 C1SC203723, mRNA sequence.
 ACCESSION AW030131
 VERSION AW030131.1 GI:5888887
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
 Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1 (bases 1 to 650)
 AUTHORS Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,
 Holt, I.E., Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S.,
 Ronning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and
 Giovannoni, J.
 TITLE Generation of ESTs from tomato callus tissue
 JOURNAL Unpublished (1999)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188364.
 Contact: David Erisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU
 5 prime sequence.
 Location/Qualifiers
 1..650
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone_lib="c1EC203723"
 /tissue_type="tomato callus, TAMU"
 /dev_stage="25-40 days old"

```
/lab_host="XL1-Blue MRF"
/notes: Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 177 a 142 c 137 g 194 t
ORIGIN
Query Match 11.6%; Score 245.2; DB 63; Length 650;
Best Local Similarity 65.3%; Pred. No. 1.7e-47;
Matches 393; Conservative 0; Mismatches 203; Indels 6; Gaps 2;
QY 310 aagcttttaattttcactaagaagaaataatgtgcagctcatgtcatcttaacaac 369
Db 24 AGTCTCTATTTTAAAGTTGAGAGTGTAATAATGTCACCCCACTGCTATTTGGTGAC 83
QY 370 atttcagcacaaagccatataatcagcagcactcaatttgcgaagaattttgcaagat 429
Db 84 ATTTTCTGCACAAAGGTCATATCAATCCATCTGTTCATTTGCCAAGAGGCTAATCGAAAT 143
QY 430 ggcataagtgagcattttctacaagcattatgcccagaagccgatgatgataaatac 489
Db 144 GGGCATGTAGGTGACATTCACCACTAGTGTTCACCCACCGTCGTATGGGAAATCGC 203
QY 490 cattottaatgcacaaagagtgaaatttccattccattttcccgatgggttgatgaag 549
Db 204 GGTTCACCGGCACCAAGGGCTTAAACTTGGCGGCATTCTCTGATGGATTGATGAG 263
QY 550 ttttga---tcattcaaaagccctgtattttacatgtcacacttcgtaaatgtggaag 606
Db 264 TTTCAGCTTACGTGGATGATCTTAAAGCTTACATGTACAGATAGAGATCGTGGGTC 323
QY 607 tgaactgtcaaaaaataattctcaactgtctgtgaaatggcagcagcctataacttgcct 666
Db 324 CCAACATTTGAGGGATGCAATTTTGAAGAGTTACAGAGGAGGAGGCTTCTGTACGTCCT 383
QY 667 actttactccatttccctcttgagcagagagtgagcagtgaaagtccactcccttc 726
Db 384 CGTCTACACTTCTGCTACCTTGGCGAGCTGATGTAGCGGGAACCTCATATCCCATC 443
QY 727 tgcctcttttggagtcacacgaacaataattggacatatatttcaacttttcattgg 786
Db 444 CGGTTTACTTGATTAACACGAACTGTCTAGACATATACTACTATTACTTCCATGG 503
QY 787 atatgaaaagctat---ggctaagaaatcaatgatccaaatgtccattcaactcc 843
Db 504 CTATGAGGATGAATGAAGTGTAGTGTCAATTCATTCATTCGAGTGTCCATGGC 563
QY 844 cgggtctccactatggaactcagatcttcccttccatttttacttccctatgggtgcaa 903
Db 564 AAGGCTTCCATTACTATAAAGCCAGATCTTCCATCTTTTTTATTGTTTCTAGTCAAA 623
QY 904 ag 905
Db 624 AG 625
RESULT 15
AW651250
LOCUS EST329704 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone cLE116E21 5', mRNA sequence.
ACCESSION AW651250
VERSION AW651250.1 GI:7412488
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
```

```
REFERENCE 1 (bases 1 to 642)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and
Tanksley,S.D.
TITLE Generation of ESTs from germinating tomato seed
JOURNAL Unpublished (2000)
COMMENT On Jan 6' 2000 this sequence version replaced gi:6677152.
Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence
FEATURES
Source
1..642
/organism="Lycopersicon esculentum"
/db_xref="TA96"
/db_xref="taxon:4081"
/clone="cLE116E21"
/clone.lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 179 a 145 c 135 g 183 t
ORIGIN
Query Match 11.3%; Score 238.4; DB 79; Length 642;
Best Local Similarity 65.8%; Pred. No. 6.6e-46;
Matches 379; Conservative 0; Mismatches 191; Indels 6; Gaps 2;
QY 310 aagcttttaattttcactaagaagaaataatgtgcagcctcatgtcatcttaacaac 369
Db 67 AGTCTCTATTTTAAAGTTGAGAGTGTAATAATGTCACCCCACTGCTATTTGGTGAC 126
QY 370 atttcagcacaaagccatataatcagcagcactcaatttgcgaagaattttgcaagat 429
Db 127 ATTTCTGCACAAAGGTCATATCAATCCATCTCTTCAATTTGCCAAGAGGCTAATCGAAAT 186
QY 430 ggcataagtgagcattttctacaagcattatgcccagaagccgatgatgataaatac 489
Db 187 GGGCATGTAGGTGACATTCACCACTAGTGTTCGCGCAGGCTGTATGGGAAATCGC 246
QY 490 cattottaatgcacaaagagtgaaatttccattccattttcccgatgggttgatgaag 549
Db 247 GGTTCACCGGCACCAAGGGCTTAAACTTGGCGGCATTCTCTGATGGATTGATGAG 306
QY 550 ttttga---tcattcaaaagccctgtattttacatgtcacacttcgtaaatgtggaag 606
Db 307 TTTCAGCTTACGTGGATGATCTTAAAGCTTACATGTACAGATAGAGATCGTGGGTC 366
QY 607 tgaactgtcaaaaaataattctcaacttgcctgtgtaaatggcagcagcctataacttgcct 666
Db 367 CCAACATTTGAGGGATGCAATTTTGAAGAGTTACAGAGGAGGAGGCTTCTGTACGTCCT 426
QY 667 actttactccatttccctcttgagcagagagtgagcagtgaaagtccactcccttc 726
Db 427 CGTCTACACTTCTGCTACCTTGGCAGCTGAGTAGCGCGTGAACCTCATATCCCATC 486
QY 727 tgcctcttttggagtcacacgaacaataattggacatatatttcaacttccattgg 786
Db 487 CGGTTTACTTGATTAACACGAACTGTCTAGACATATACTACTATTACTTCAATGG 546
QY 787 atatgaaaagctat---ggctaagaaatcaatgatccaaatgtgccattcaactcc 843
Db 547 CTATGAGGATGAATGAAGTGTAGTAGTCAATTCATTCATTCGAGTATCCAAATGGC 606
```

Qy 844 cgggcttccactactggaaactcgagatcttcttc 879
||||||| ||| | ||||| ||
Db 607 AAGGCTTCCATTACTAAAAAGCCAGATCTTCATC 642

Search completed: August 1, 2000, 14:38:55
Job time: 62727 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2000, 21:27:45 ; Search time 51.03 Seconds
(without alignments)
559.061 Million cell updates/sec

Title: US-09-147-955-6
Perfect score: 2407
Sequence: 1 MSRAHVLATFPAQGHINPA.....EDGSSVNNLKVFLDEVVGI 461

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-64: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1054.5	43.8	456	2 C71402	probable glucosylit
2	1052	43.7	467	2 T02238	glucosyl transfera
3	935.5	38.9	458	2 C71420	hypothetical prote
4	751.5	31.2	471	2 A54739	indole-3-acetate b
5	740	30.7	460	2 T00639	hypothetical prote
6	711.5	29.6	479	2 E71419	probable indole-3-
7	697	29.0	456	2 T00506	indole-3-acetate b
8	682	28.3	438	2 T00507	indole-3-acetate b
9	656	27.3	453	2 D71419	probable indole-3-
10	631	26.2	475	2 T00511	probable indole-3-
11	631	26.2	475	2 F71419	probable indole-3-
12	504.5	21.0	455	2 T00584	indole-3-acetate b
13	483	20.1	440	2 T00583	indole-3-acetate b
14	482	20.0	453	2 T45603	glucosyltransferas
15	479	19.9	462	2 T01732	UTP-glucose glucos
16	457	19.0	476	2 T03745	glucosyltransferas
17	445	18.5	476	2 T03747	glucosyltransferas
18	443	18.4	451	2 T45604	glucosyltransferas
19	441	18.3	507	2 T46161	glucosyltransferas
20	431	17.9	347	2 T06371	probable UDP-glucu
21	429.5	17.8	433	2 S51767	glycosyl transfera
22	429	17.8	466	2 T07404	probable glucosylit
23	428	17.8	478	2 T05423	probable glucosylit
24	426.5	17.7	490	2 T46162	glucosyltransferas
25	424.5	17.6	449	2 T45605	glucosyltransferas
26	419.5	17.4	452	2 T12981	hypothetical prote
27	416.5	17.3	487	2 S41951	UTP-glucose glucos
28	412.5	17.1	438	2 T45602	glucosyltransferas
29	405.5	16.8	488	2 T07786	UDP-glucose glucos

30	399	16.6	447	2 T12978	hypothetical prote
31	393	16.3	455	1 XUBHFG	flavonol 3-O-gluco
32	389	16.2	452	2 T00981	flavonol 3-O-gluco
33	389	16.2	481	2 T01850	UTP-glucose glucos
34	386	16.0	478	2 A71417	hypothetical prote
35	383.5	15.9	478	2 T08395	UTP-glucose glucos
36	383	15.9	385	2 T12980	hypothetical prote
37	382	15.9	471	2 S01052	flavonol 3-O-gluco
38	379	15.7	420	2 T08005	flavonol 3-O-gluco
39	372.5	15.5	452	2 G71416	probable glucosylit
40	371.5	15.4	346	2 S41954	UTP-glucose glucos
41	366	15.2	471	2 S01037	flavonol 3-O-gluco
42	363.5	15.1	287	2 S41953	UTP-glucose glucos
43	362.5	15.1	472	2 S39507	glucuronosyl trans
44	361	15.0	449	2 S41950	UTP-glucose glucos
45	360	15.0	471	1 S08325	flavonol 3-O-gluco

ALIGNMENTS

RESULT 1

C71402
Probable glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: C71402
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoff, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113
A:Accession: C71402
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-456 <BEV>
A:Cross-references: GB:I297335; NID:92244747; PID:e3274440; PID:g2244766
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 43.8%; Score 1054.5; DB 2; Length 456;
Best Local Similarity 49.7%; Pred. No. 2.7e-73;
Matches 232; Conservative 70; Mismatches 130; Indels 35; Gaps 14;

Qy	3	RAHVLLATFPAQGHINPALQFAKRLANADIQVTFETSYVAMRRMSRTAAGSNGLINEVSVF	62
Db	11	RPHVLLTFFPAQGHINPALQLANRLTHGATVTYTSVSAHRRMGEPFS-TKGL-SFAWF	68
Qy	63	SDGYDDGLQPGDDGKNYSEKSRGKALKSDTLAAN-NVDQKSKITFYVYSHLFAWAAK	121
Db	69	TGDEDDGLKSFQDKIYSEKSRGNALNDITKANLDATTEPTGYIVSVLPVWYST	128
Qy	122	VAREFHLSALLTETATVLDIFYYFN-GYSEIDAGSDAHLPGGLPVLQAQRLPSFL	180
Db	129	VAREFHLPPTLLTETATVLDIVYVFNSTYKHLFDV--EPIKLP-KLPLITGDLPSFL	185
Qy	181	LPSTHEFRSL-----MKEKLEGEKPKVNVNSPDALPDALKADKVMETAGPL	234
Db	186	QPS-----KALPSALVTLREHIALETESPKLVNFTSALEHDLTSEVKLMTPIGFL	240
Qy	235	IPSAFLDGKOPSDRSGDGLFEKGSNDQCLEWLNTPRSPVYVSVFGSFV-NTTKSQME	293
Db	241	VSSS-EGKT-----DLFK--SSDEDYTKWLSKLSRSVIYISLGTADDLPEKME	288
Qy	294	ETANGLLDCGPPFLWVVRVNEGEVLISCHEELKR---VGKIVSNCSQVLETHPSLGF	350

Db 289 ALTHGVLATNRPFLWVBEKPKKRNFLIRGSDRGLVVGWCSQTAFLAHCAVGC 348
 QY 351 VTHCGWNTLSISFGVPMVAFPOWFOGNTAKLMEDVWRTGVRVRANBEGSVVGDDEIR 410
 Db 349 VTHCGWNTLSISFGVPMVAFPOWFOGNTAKLMEDVWRTGVRVRANBEGSVVGDDEIR 407

QY 411 RCIEEVMDDGGKSRKLRSAKWKDLARKAMEEDGSSVNNLKVFLDE 457

Db 408 RCLKVMGSGEAEEMRENAEKWAMVDAEAEGGSDNLKGFVDE 454

RESULT 2

T02238
 C:Species: Nicotiana tabacum (common tobacco)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 24-Nov-1999
 C:Accession: T02238
 R:Kojima, H.; Hashizume, K.; Imanishi, S.; Nakamura, K.
 submitted to the EMBL Data Library, January 1997
 A:Description: Jasmonate-induced potential glucosyltransferase from tobacco suspension
 A:Reference number: Z14633
 A:Accession: T02238
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-467 <KQ>
 A:Cross-references: EMBL:AB000623; NID:d1094897; PID:d1019901
 A:Experimental source: strain BY-2
 C:Genetics:
 A:Gene: JIGT
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match 43.78; Score 1052; DB 2; Length 467;
 Best Local Similarity 45.11; Pred. No. 4.3e-73;
 Matches 211; Conservative 93; Mismatches 140; Indels 24; Gaps 11;

QY 2 SRAHLLATFFPAQGHINPALQFAKRLANADIQVTFSTVYAMRMRSTAAAGSNGLINEFS 61
 Db 6 NECHVLIALLFPQGHINPSLQFSKRLNLGVKVTLSLSLAFNRI-KNLPKIEGL-TPAP 63

QY 62 FSDGVDGDLQPG-DDGKNYEMSEKSRGKALSDTLAANNVDQSSK---ITFVYVSHLFA 117
 Db 64 FSDGVDGDLQPG-DDGKNYEMSEKSRGKALSDTLAANNVDQSSK---ITFVYVSHLFA 117

QY 118 WAQVAREFHRLSALLWTEPATVLDIFVYFNGYSDEI---DAGSDAHLPGGLPVAQR 174
 Db 119 WAGVAKKLHIPSTLFWIQPATVFDIYVYFTNFANFYKDYSDQDIIELP-GLPSLSSS 177

QY 175 DLPSFLPSTHERPRL--MKEKLETEGEEKPKVLVNSFDALPDALKADIKYEMIAIG 232
 Db 178 DFPSEFDDVKSNDWAVESIKRQIEILNSENPRILYNTDALELNALRLKNTVAVGIC 237

QY 233 PLIPSAFLDGKDPDRSGGDLFEKGSNDODCLEWLSNPRSSVYVYVSGSFVNTTKSQM 292
 Db 238 PLIPSEFLDEKDRKNDFADNIE---SENNYMEWLDAKANSVIYIAFGVSAEISSQNM 294

QY 293 EBIAGLLDCGRLPFLWVVRV--NEGSEVLISCMEELKRVGKIVSCSOLEVLTHPSLGC 349
 Db 295 EBIISQGLLKCGRPLFWITRETLNKEKPEEKLTKDELEKIGRIVRWCSONEVLKHSVGC 354

QY 350 FVTHCGWNTLSISFGVPMVAFPOWFOGNTAKLMEDVWRTGVRVRANBEGSVVGDDEIR 409
 Db 355 FLTHCGWNTLSISFGVPMVAFPOWFOGNTAKLMEDVWRTGVRVRANBEGSVVGDDEIR 413

QY 410 RRCIEEVMDDGGKSRKLRSAKWKDLARKAMEEDGSSVNNLKVFLDE 457
 Db 414 QKCIENVGDAEAGEELKRAKQKWLKASTKENSNNVNLKAYNE 461

RESULT 3
 C71420
 hypothetical protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: Columbia
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
 C:Accession: C71420
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terryn, N.; G
 avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
 erioft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
 C.; Chalwatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
 A:Reference number: A71400; MUID:98121113
 A:Accession: C71420
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-458 <BE>
 A:Cross-references: GB:Z97339; NID:G2244901; PID:G326935; PID:G2244911
 C:Genetics:
 A:Map position: 4COP9-4G3845
 C:Superfamily: flavonol O3-glucosyltransferase

Query Match 38.9%; Score 935.5; DB 2; Length 458;
 Best Local Similarity 42.78; Pred. No. 3.7e-64;
 Matches 206; Conservative 82; Mismatches 126; Indels 69; Gaps 16;

QY 5 HVLLATFFPAQGHINPALQFAKRLAN--ADIQVTFSTVYAW-RMRSTAAAGSNGLINEFS 61
 Db 13 HFLVTFPAQGHINPSLELAKRLAGTISGARVTPAASISAYNRMPSTENVPETLI-FAT 71

QY 62 FSDGVDGDLQPG-----DDGKNYEMSEKSRGKALSDTLAANNVDQSSKITFVYVS 113
 Db 72 YSDGHDGKFSAYSDKSRQDAGTGNFSEMRRRKGTETELIEDNR--KQNRFTCVVYT 129

QY 114 HLFWAQVAREFHRLSALLWTEPATVLDIFVYFNGYSDEI---DAGSDAHLPGGLP 169
 Db 130 ILLTWVAELA-----LFSIFVHYFNGYDAISEMANTPSSSKLP-SLP 172

QY 170 VLAORDLPFLPSTHERP-RLMKLETEGEEKPKVLVNSFDALPDALKAI-DKYE 227
 Db 173 LLTVRDIPEFIVSSNVYAFLLPAFREQIDSLKEINPKILINTFOELEPKMSVDPNPK 232

QY 228 MIAIGPLIPSAFLDGKDPDRSGGDLFEKGSNDODCLEWLSNPRSSVYVYVSGSFVNT 287
 Db 233 IVPVGPGL-----TLRTDFSSRG---EYEWLDTKADSSVLYVSGFLAVL 275

QY 288 TKQMEIEIARGLLDCGRPLWV-----RVNGESEVLISCMEELKRVGKIVSWCS 337
 Db 276 SKQLVELCKALIQSRPFLAVITDKSYRNKEDEQEEDCISSEFLEDEIGWVSWCD 335

QY 338 QLEVLTHPSLGCFTVTHCGWNTLSISFGVPMVAFPOWFOGNTAKLMEDVWRTGVRV-- 395
 Db 336 QFRVLNHRISIGCFVTHCGWNTLSISVSGVPMVAPPQNMNDMMNAKLIEDCWTGTVRME 395

QY 396 RANEEG-SVVDGDEIRRCIEEVMDDGGKSRKLRSAKWKDLARKAMEEDGSSVNNLKV 454
 Db 396 KKEEGGVVVVDSEIRRCIEEVM--DKAEFRGNATRWDLAAEAAREGSSFNILKAP 453

QY 455 LDE 457
 Db 454 VDE 456

RESULT 4
 A54739
 indole-3-acetate beta-glucosyltransferase (EC 2.4.1.121) - maize

C:Species: Zea mays (maize)
 C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 29-Oct-1999
 C:Accession: A54739
 R:Szerszen, J.B.; Szczyglowski, K.; Bandurski, R.S.
 Science 265, 1699-1701, 1994
 A:Title: Iaglu, a gene from Zea mays involved in conjugation of growth hormone indole
 A:Reference number: A54739; MUID:94367368

A:Accession: A54739
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-471 <SIZE>
A:Cross-references: GB:134847; NID:g548194; PID:AAA59054.1; PIR:1000000000
C:Superfamily: flavonol O3-glucosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match	31.28;	Score	751.5;	DB	2;	Length	471;
Best Local Similarity	38.18;	Pred.	No. 5.1e-	50;			
Matches	182;	Conservative	75;	Mismatches	180;	Indels	41;
	Gaps	13;					
QY	5	HVLLATFPQAQGHINPALQFAKRLANADIQVTFFTSVYAWRMKRTAAGSNGLINFVSFSD	64				
Db	4	HVLVFPFGQGHNNPWFQFAKRLASKGVATTLVTT----	59				
QY	65	GYDDGLQPGDDCKNTYMKSEKRGTKALSDTLAAANNVDOKSSK----	121				
Db	60	GHDEG----GFASAAGVAYELKQAAAAASL-ASLVEARASSADAFTCVVYDSYEDWVLP	115				
QY	122	VAREHLRSGALLWIEPATVLDIFYFYFNG-----YSDEIDAGSDAHLPG----	170				
Db	116	VARMGLPAVPSTQSCAVSYVTHFSQRLVPPGAANDGSGGAGAAALSAFGLDPE	175				
QY	171	LAORDLPFLPSTHERFSRLMKKELETLEGEERPK-VLVNSFDALEPDALKAIKY-EM	228				
Db	176	MERSELPSEVF--DHGPYTTAMQAIKQFAHAGKDDWLVFNSPEELETEVLAGTLKYLKA	233				
QY	229	IATGPLIP-----SAFLDGKDPDSRFGGDLPEKSGNDDDCLEWLSLTPSRSSVYVYSG	282				
Db	234	RALGPCVPLPTAGRTAGANGR-----ITYGALNLKP-----EOACTKWLDTKPDRSVAIVSG	286				
QY	283	SFYNTTTSQMEIARGLLDCGRPFLVYVVRNVEGEVLISCMIELKRVG-KIVSMCSQLE	340				
Db	287	SLASLNAQKEELARGLLAAGKPFLLVYVVRASDEHOVPYRLAEATATGAAYVPMCPQLD	346				
QY	341	VLTHPSLGCFTVTHCGWNSTLESI SRGPMVAFFQWFDGNTNAKLMEDEVWRTGVYVRANEE	400				
Db	347	VLAHPAVGCFTVTHCGWNSTLEASLPGVPMWAMALWTDQPTNARNVELAMGAGVRRADAG	406				
QY	401	GSVVDGEIIRRCIEEYMDGGEKSRKLRESAGKWKDLARKAMEEDGGSYNNLKVFLDEV	458				
Db	407	AGVFLRGEYERCVRAVMDGGEAASAKAAGEWDRARAAVAGPGSSDRNLNDFGVQV	464				

RESULT 5
T00639
hypothetical protein F316.2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C:Accession: T00639
R:Pederspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,
J.; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z14197
A:Accession: T00639
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-460 <FED>
A:Cross-references: EMBL:AC002396; NID:g2745918; PID:g2829862; GSPDB:GN00059; ATSP:F316.2
C:Genetics:
A:Gene: ATSP:F316.2
A:Map position: 1
A:Introns: 219/3
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 30.7%; Score 740; DB 2; Length 460;
Best Local Similarity 36.5%; Pred. No. 3.8e-49;
Matches 173; Conservative 78; Mismatches 181; Indels 42; Gaps 13;
QY 3 RAHVLLATPPAGGHINPALQFAKRLANADIQVTFETTSVYAWRRMRS TAAG-SNGLINFVS 61

Db 9 KGHVITLPPVOGHLNPMWOFKRLSKNVKVTIATTV-----TASSTTTPSLSVGP 61
QY 62 FSDGYD--DGLQPCDCKNYMSEKSRGFKALSUDTFLAANNVDOKSKKITFVVYSHLFAW 118
Db 62 ISDGFDPFIGI-PCGFSVDYTVSEFSLKNGSETL--TLLIEKFKSTDSPIDCLLYDSFLPW 118
QY 119 AKVAREPHILSALLWTEPATYLDIFTFYFNGYSDEIDAGSDAIHLP---GGLPVLQAORD 175
Db 119 GLEVARSMELSAASFNNLTNYSVURKFSFG---DFPLADPNAPSAPRIRGLPSLSYDE 175
QY 176 LPSFSLIPS--THERFSRLMKLEKLETEGEKPKVLVNSPDAL-----EPDALKKAI 224
Db 176 LPSFVGRHWLTHPEHGRVLLNQPNHENAD--WLPVNGPEGLEETQDCENGESDAMKA-- 231
QY 225 KYEMTAIGPLTPSAPLOGKDDPSRSGGDLFEKGSNDDCLLEWLSNTNPRSSVYTVSGSF 284
Db 232 ---TLIGFMPITSAIYLDNRMDKDYGASLLPKISK--ECMEWLETKQAQVAFVSGSF 285
QY 285 VNNTKSOMEETARGLLDCGRFPLVVRVNGEVEVLITSCMEELKRVGKITVSCMSOLEVLTH 344
Db 286 GILFEKQLAEVAIALOESDLNFWLWIKEAHIAKLPGFVSTKORALLVSHCNQLEVLAH 345
QY 345 PSLGCFVTHCGNNSTLESISGVPWMAFPQWFDQGTNAKLMEDVMRTGVRVRANEESGVV 404
Db 346 ESIGCFLTHCGNNSTLEGLSLGVPMVGPQWSDQMDAKAFVEEVVKVYRAKEAGEVIV 405
QY 405 DGDTRCITEEVMDGGKSRKLESAGKWKDLARKAWEDGGSSVNLKVLDEV 458
Db 406 KSEELVRCILKGVME--GESSVKIRSSKKKDLAVKANSBGSDSKRINEFISL 458
RESULT 6
E71419 probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-car cross)
A:Variety: Columbia
C:date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 24-Nov-1999
C:Accession: E71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Berg
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Te
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer,
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Re
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Ar
A:Reference number: A71400; MUID:98121113
A:Accession: E71419
A:Status: preliminary; nucleic acid sequence not shown; translation not sh
A:Molecule type: DNA
A:Residues: 1-479 <BE>
A:Cross-references: GB:297339; NID:g22444901; PID:e326931; PID:g22444906
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 29.6%; Score 711.5; DB 2; Length 479;
Best Local Similarity 35.8%; Pred. No. 6.2e-47;
Matches 172; Conservative 89; Mismatches 174; Indels 45; Gaps 15;

```
Qy      5 HVLLATFPAQGHINPALOFAKRILANADIOVTFTSVYARRMSRTAAGSN-----G
Db      8 HVMLSFFPGQHVNPLRLJOKLITASGLLVLTFTVTEKPMGWKQANKIQDGLKPVGLG 67
Qy     56 LINFVSFDGDYDGLQGPDGGKNYMSEMSRGIKALSITDLAANNVDQSSK-ITFYVYSH 114
Db     68 FIRFEFFSDGAD----DDEKRFDFAFRPLEHVAVGQEIKNLVKRYNEPVTCLINNA 122
Qy    115 LPAAWKAARVFPHLRSAALLMIEPATVLIDIFYEFNGY-----SDEIDAGSDATHLPGLP 169
Db    123 EYPWCWDVAELHLPISVALWOSACLTAIVYYHHRIKVEFKTEPDI-----SWEIP-CUP 178
```

[illegible]

Qy	131	ALLWIEPATVLDIVFYFYP--NGYSDIEIDAGSDAHLHGGGLPVLAQRDLPSFLLPSTH	18
Db	115	AIUHQACAGFSYTRYTKNTPPDLEU-NGTVELP-ALPILLEVRDLPSLUMPOGAN	172
Qy	188	FRSLMKELLETLEGEEKPKVLVNSFDALPDALKAI-DKYEMIAIGPLIPSAFIDGKDQPS	246
Db	173	VNTLMAEPADCL--KDKWVLVNSFYELSEIESNDLKPIIPIGPLV-SPFLLGNDDE	229
Qy	247	DRSFGGDLPEKSGNDDDCLEWLSNTPRSSVVVYVSGFSVYNTTKSQMBEIRAGLLDQGRPF	306
Db	230	-----KTLDMMKVDDYCMEWLDQKQARSSVVYISFGSILKSLNQVETIATALKNRGVPP	283
Qy	307	LWVVRVNE-GEVVLISCHEELKRVKG--IVSWCSQLEVLTHPSLGCFVTHCGNNTLESI	363
Db	284	LWVIRPKKGENVQV--LQEMVKEGKGVVTMEQQEKILSHMAISCFITHCGMNSTIETV	341
Qy	364	SEGVPVAPQFQDGTNAKLMEVDVHRTGVVRVRADEGSVVDGD-----ETRRICEEYMDG	419
Db	342	VTGVPVVAFTWIDQPLDARLLVDVFGIGVRMK---NDAIDGELKVAEVRCEIAVTE-	396
Qy	420	GEKSRKLRESACKWDLARKAMEDEGSSVNNKLKVFIDEV	458
Db	397	GPAADMRRRATRELKHAARSAMSPGGSSAQNLDSFTSDI	435

RESULT 9
 D71419
 probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia
 C:Date: 03-Aug-1998 #sequence revision 03-Aug-1998 #text change 24-Nov-1999

C:Genetics:
A:Map position: 2
A:Introns: 266/2
A:Note: T20D16.16
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 27.3%; Score 656; DB 2; Length 453;
Best Local Similarity 34.9%; Pred. No. 1e-42;
Matches 164; Conservative 88; Mismatches 164; Indels 54; Gaps 18;

QY 5 HVLFTTAPQAQHINPQLQFAKRLANADIQVFTFTSVYAWRRSRTAAGSNGLINFYFSFD 64
DB 10 HVLWVALPFGQHLNPMKPKAKHLARTNHTLAT-IESARDLLSSTDPSHSLVDLVFFSD 68
QY 65 GYDDGLPGDDGKNYSEMKSRIKALSDTL---AANNVQ--KSKITFVYVSHLFAWA 119
DB 69 GL-----PKDDPRDH-----ELTSLRKVGANFSKIIEGRKFCIIISVPTTPW 114
QY 120 AKVAREFHLSALLWIEPATVLDIFYFYF---NGYSEIDAGSDAHLPGGLPVLAQRDL 176
DB 111 PAVAAAHNIPCAILWIEACAGFSVYRYRYMKTNSFPD-LEDPNQKVELP-GLPFLSVROL 172
QY 177 PSLPLSTHERFRSLMKELETLEGKPKVLVNSDALEPDALKAI-DKYEMIAIGPLI 235
DB 173 PTLMPSHGAIFTLMAEFVECL--KDKVVLANSYELESVIESWFDLKPPIIGPLV 230
QY 236 PSAPLDGKPSDRSFGDGLFEKSGNDDCLEWLSNTNPRSVYVSGSVFNTTKSQMERI 295
DB 231 -SPFLLAGD-EDKILDGKSLDMKADDCYCMWLDKQV-----SILKSSNOVETI 278
QY 296 ARGLLDGRFLAVWRVNEGE--VLISCMELKRVGKIVSVCSOLEVLTHPSLGCFVT 352
DB 279 ATALKNRGVFLVIRPKAEVNDVLEDWVEGQ--GVVIEGQGEKILCHWAISCFVT 336
QY 353 HCGNSTLESISFGVPVAPFPQNDGTNAKLMDVNRGTQVVRANESVSDG-----E 408
DB 337 HCGWNSITETVSGVPVAVPTWFDPLDARLLVDVFGVGRMK----NDVVDGELKVAE 392
QY 409 IRRICIEVMDGGSKRKRESACKWKDLARKAMEEDGSSVNNLKVFLDVE 458
DB 393 VERCIDAVTGTG-DADMRRAELKQATRSAMAPGSLARNLDLFINDI 441

RESULT 11
F71419
probable indole-3-acetate beta-glucosyltransferase - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: columbia
C:Accession: F71419
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D.
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; G.
anavath, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chludzisz, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113
A:Accession: F71419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <BEV>
A:Cross-references: GB:297339; NID:g2244901; PID:e3226932; PID:g2244907
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 26.2%; Score 631; DB 2; Length 475;
Best Local Similarity 33.3%; Pred. No. 9.2e-41;
Matches 165; Conservative 87; Mismatches 173; Indels 70; Gaps 19;

[illegible]

RESULT I3
T00593

indole-3-acetate beta-glucosyltransferase homolog T27E13.11 - Arabidopsis
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 11-Jan-2000
C:Accession: T00583

RESUL I 2
T00584

A:Accession: T00583
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-440 <R0U>
A:Cross-references: EMBL;AC004165; NID:g3150396; PID:g3150405
C:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 149/1
A:Note: T27EL3.11
C:Superfamily: flavonol O3-glucosyltransferase

Query Match 20 18. Score 483. Dg 2. Length 440.

Query Match 20.1%; Score 483; DB 2; Length 440;
Best local similarity 29.8%; Pred No. 1 9e-29.

Best Local Similarity 23.6%; Pled. NO. 1.9e-29;
Matches 141: Conservative 83: Mismatches 191: Indels

Matchless 141, conservative 93, misadventures 131, riders

0Y 11 FPAOCHINPALOFAKRLANADIOVTF-FTSVYAWRRMSRTAAGSN--GLINFV

[illegible]

Query Match	21.0%	Score	504.5	DB 2:	Length	455.
-------------	-------	-------	-------	-------	--------	------

59 PSEI V B A N D E I A F I D A V I T R I E F F E E O I ! - - - - - D R I N S P P T A I A C T V I I W A V

DB 39 PSELVRANDFIAFIDAVLTRLRLEEPFEQLL-----DRLNSPPTAIIADTYIIWAV

QY 127 H1.B.SA.L1.WTE.PAT.VI.DIE-----VEFENGYSDEI.DAGSDA.THI.PGGI.BV

QY 127 HLRSALEWIEPATVLDIF-----IFYNGISDEIDAGSDATHLPFGGLPY

114 NIPVASEFWTTSATILSLEFNSDILLASHGHFPIEPSESKLDETVD--YIPGI.SPT

114 NIT VABE WIT 304 T E S E I N S D E L A R S H O U F F I E F S E S N E D D I A V D - I A F O U S F I

QV 177 PSELLPSTHERFRSLMKEKLETLEGEKPKVLNSFDALPDALKAD-----

[illegible]

Db 171 -QILHGYSHQVFNIFKKSGELYKAK---YLLFPSAYELEP---KAIDFTSKF

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

QY 231 IGPLIPSAFLDGCKDPDRSFGGDLFEKGSNDDDCLEWLSTNPRSSVVYVSGSF

[illegible][illegible]

Search completed: August 1, 2000, 21:27:46

Fri Aug 4 15:19:52 2000

us-09-147-955-6.rpr

Page 8

Job time: 11214 sec